Assesssment 1 - COVID-19 impact on digital learning

Nguyen Quoc Trung / Student ID: 48299405

2024-04-12

During the COVID - 19 pandemic, more than 56 million students in the United witnessed the interruption in their education. In the spring of 2020, the US government decided to shut educational institutes to stop the virus from widely spreading. Schools and teachers started teaching students remotely by adapting several online tools and digital platforms. In other words, online teaching became a necessity and widely adopted by educational institutions during the lockdown. Until today, concerns of the exacerbating digital divide and long-term learning loss among America's most vulnerable learners continue to grow.

This report attempts to explore digital learning trends by analyzing a dataset from various U.S. districts to e and how students use different tools and platforms. To be more specify, the relationship between engagement in digital learning for different states will be conducted to work on this matter.

1. Data Cleaning and Wrangling

a. Importing data files:

In the very first stage, this report is working on importing the different datasets into the data frame. The construction of each files is following on the below observation on the raw data file:. Firstly, the districts data file is imported to the working workspace.

```
districts <- read_csv(file = 'districts_info.csv') # import the district file
to working space in R_studio

## Rows: 233 Columns: 7
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr (6): state, locale, pct_black/hispanic, pct_free/reduced,
county_connect...
## dbl (1): district_id
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

head(districts) # See the first few lines of the districts file

## # A tibble: 6 × 7
##   district_id state     locale `pct_black/hispanic` `pct_free/reduced`
##         <dbl> <chr>     <chr>  <chr>                <chr>
## 1        8815 Illinois  Suburb [0, 0.2[             [0, 0.2[
```

```
## 2        2685 NaN       NaN    NaN                   NaN
## 3        4921 UTAH      Suburb [0, 0.2[              [0.2, 0.4[
## 4        3188 NaN       NaN    NaN                   NaN
## 5        2238 NaN       NaN    NaN                   NaN
## 6        5987 Wisconsin Suburb [0, 0.2[              [0, 0.2[
## # ℹ 2 more variables: county_connections_ratio <chr>, pp_total_raw <chr>
```

```
colSums(is.na(districts)) # check the NA for each column in the data file
```

```
##              district_id                    state                   locale
##                        0                        0                        0
##       pct_black/hispanic         pct_free/reduced county_connections_ratio
##                        1                       28                       14
##             pp_total_raw
##                       58
```

```
knitr::opts_chunk$set(echo = TRUE)
```

Then, the products file is imported.

```
products <- read_csv(file = "products_info.csv") # import the products file
```

```
## Rows: 372 Columns: 6
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (5): URL, Product Name, Provider/Company Name, Sector(s), Primary Essent...
## dbl (1): LP ID
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(products) # See the first few lines of the products file
```

```
## # A tibble: 6 × 6
##   `LP ID` URL                  `Product Name` Provider/Company Nam…¹ `Sector(s)`
##     <dbl> <chr>                <chr>          <chr>                  <chr>
## 1   13117 https://www.splashm… SplashLearn    StudyPad Inc.          PreK-12
## 2   66933 https://abcmouse.com ABCmouse.com   Age of Learning, Inc   PreK-12
## 3   50479 https://www.abcya.c… ABCya!         ABCya.com, LLC         PreK-12
## 4   92993 http://www.aleks.co… ALEKS          McGraw-Hill PreK-12    PreK-12; H…
## 5   73104 https://www.achieve… Achieve3000    Achieve3000            PreK-12
## 6   37600 http://www.actively… Actively Learn Actively Learn        pre kinder…
```

```
## # ℹ abbreviated name: ¹`Provider/Company Name`
## # ℹ 1 more variable: `Primary Essential Function` <chr>
```

```
colSums(is.na(products)) # check the NA for each column in the data file
```

```
##                       LP ID                          URL
##                           0                            0
##                Product Name      Provider/Company Name
##                           0                            1
##                   Sector(s) Primary Essential Function
##                          18                           20
```

```
# After review, the products file needs to change the column name from LP_ID
to lp_id to match the format used in five district files within the
engagement data. This reformatting is crucial for merging the files
correctly.
```

```
names(products)[names(products) == "LP ID"] <- "lp_id"
```

```
knitr::opts_chunk$set(echo = TRUE)
```

Lastly in this stage, each file in the folder represents data from individual schools, and the engagement data are aggregated at the school district level will be imported. Moreover, it has been observed that for effective merging of datasets, a 'district_id' column must be added to each school's data file. This addition provides a corresponding district identifier for each school, ensuring that data can be accurately aligned and combined across the district level.

```
df1000 <- read_csv(file = "1000.csv") # Import data file district 1000
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame
for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 104003 Columns: 4
## ── Column specification
──────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): time
## dbl (3): lp_id, pct_access, engagement_index
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

```
df1000$lp_id <- as.numeric(df1000$lp_id) # change lp_id
engage1 <- mutate(df1000, district_id = 1000) # add the district_identifier
engage1 # re name the data file  to engage 1
```

```
## # A tibble: 104,003 × 5
##    time      lp_id pct_access engagement_index district_id
##    <chr>     <dbl>      <dbl>            <dbl>       <dbl>
##  1 1/01/2020 93690       0                NA          1000
##  2 1/01/2020 17941       0.03              0.9        1000
##  3 1/01/2020 65358       0.03              1.2        1000
##  4 1/01/2020 98265       0.57             37.8        1000
##  5 1/01/2020 59257       0                NA          1000
##  6 1/01/2020 90153       0.06              3.9        1000
##  7 1/01/2020 41587       0                NA          1000
##  8 1/01/2022 29322       0.06              5.1        1000
##  9 1/01/2020 37479       0                NA          1000
## 10 1/01/2020 51340       0.09              1.2        1000
## # ℹ 103,993 more rows
```

```
df1039 <- read_csv(file = "1039.csv") # Import data file district 1039
```

```
## Rows: 38791 Columns: 4
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr (2): time, lp_id
## dbl (2): pct_access, engagement_index
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

```
df1039$lp_id <- as.numeric(df1039$lp_id) # change lp_id to numerical:
```

```
## Warning: NAs introduced by coercion
```

```
engage2 <- mutate(df1039, district_id = 1039) # add the district_identifier
engage2 # re name the data file to engage 2
```

```
## # A tibble: 38,791 × 5
##    time     lp_id pct_access engagement_index district_id
##    <chr>    <dbl>      <dbl>            <dbl>       <dbl>
##  1 1/1/2020 64947       0.09             0.91        1039
##  2 1/1/2020 52758       0.09             1.83        1039
##  3 1/1/2020 25559       0.09            13.7         1039
##  4 1/1/2020 13496       0.09             4.57        1039
##  5 1/1/2020 94984       0.09             1.83        1039
##  6 1/1/2020 95731       0.91            59.4         1039
##  7 1/1/2020 80616       0.27             6.4         1039
##  8 1/1/2044 47364       0.18            10.0         1039
##  9 1/1/2020 57949       0.09             1.83        1039
## 10 1/1/2020 68019       0.09             3.66        1039
## # ℹ 38,781 more rows
```

```
df1044 <- read_csv(file = "1044.csv") # Import data file district 1044
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame
for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Rows: 255722 Columns: 4
## ── Column specification
──────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): time
## dbl (3): lp_id, pct_access, engagement_index
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

```
df1044$lp_id <- as.numeric(df1044$lp_id) # change lp_id
engage3 <- mutate(df1044, district_id = 1044) # add the district_identifier
engage3 # re name the data file  to engage 3
```

```
## # A tibble: 255,722 × 5
##    time      lp_id pct_access engagement_index district_id
##    <chr>     <dbl>      <dbl>            <dbl>       <dbl>
##  1 1/01/2020 99792       0.02             2.26        1044
##  2 1/01/2020 80493       0.01             0.17        1044
##  3 1/01/2020 16164       0.03             0.35        1044
##  4 1/01/2020 61384       0.01            16.5         1044
##  5 1/01/2020 88925       0.02             0.17        1044
##  6 1/01/2020 30114       0.01             0.35        1044
##  7 1/01/2020 57293       0.01             0.09        1044
##  8 1/01/2020 85603       0.01             0.52        1044
##  9 1/01/2020 83248       0.01             0.26        1044
## 10 1/01/2020 52409       0.01             0.7         1044
## # ℹ 255,712 more rows
```

```
df1052 <- read_csv(file = "1052.csv") # Import data file district 1052
```

```
## Rows: 91977 Columns: 4
## ── Column specification
──────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): time
## dbl (3): lp_id, pct_access, engagement_index
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

```r
df1052$lp_id <- as.numeric(df1052$lp_id) # change lp_id
engage4 <- mutate(df1052, district_id = 1052) # add the district_identifier
engage4 # re name the data file  to engage 4
```

```
## # A tibble: 91,977 × 5
##    time      lp_id pct_access engagement_index district_id
##    <chr>     <dbl>      <dbl>            <dbl>       <dbl>
##  1 1/01/2020 57513       0.49           141.          1052
##  2 1/01/2020 42091       0.04             0.41        1052
##  3 1/01/2020 53399       0.04             8.19        1052
##  4 1/01/2020 32213       0.9             44.2         1052
##  5 1/01/2020 85836       0.04             0.41        1052
##  6 1/01/2033 94158       0.04             0.41        1052
##  7 1/01/2020 26488       0.41           118.          1052
##  8 1/01/2020 99916       1.19            80.7         1052
##  9 1/01/2020 92867       0.04             2.05        1052
## 10 1/01/2020 16164       0.04             0.41        1052
## # ℹ 91,967 more rows
```

```r
df1131 <- read_csv(file = "1131.csv") # Import data file district 1131
```

```
## Rows: 32963 Columns: 4
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): time
## dbl (3): lp_id, pct_access, engagement_index
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

```r
df1131$lp_id <- as.numeric(df1131$lp_id) # change lp_id
engage5 <- mutate(df1131, district_id = 1131) # add the district_identifier
engage5 # re name the data file  to engage 5
```

```
## # A tibble: 32,963 × 5
##    time      lp_id pct_access engagement_index district_id
##    <chr>     <dbl>      <dbl>            <dbl>       <dbl>
##  1 1/01/2020 16753       0.04             3.58        1131
##  2 1/01/2020 80560       0.09             8.94        1131
##  3 1/01/2020 85682       0               NA           1131
##  4 1/01/2020 95731       0.09             2.24        1131
##  5 1/01/2020 28504       0               NA           1131
##  6 1/01/2020 57084       0.04             0.89        1131
##  7 1/01/2020 25298       0.04             0.45        1131
##  8 1/01/2020 32213       0.04             0.45        1131
##  9 1/01/2020 50479       0.04             0.45        1131
## 10 1/01/2020 96111       0               NA           1131
## # ℹ 32,953 more rows
```

```
knitr::opts_chunk$set(echo = TRUE)
```

After importing the whole datafiles, it can be seen that there are 230 districts across US in this data set in 2020. Moreover, it is reported that there are 372 most used products or tools for digital learning. It can be seen that products and districts files has a missing values (values). Specifically, for the districtts files, there are several values are marked with “NaN” suggesting that the annoymization of the dataset is maximized.

b. Merging data:

In the next stage of this report, we will first merge all school files from each district in the engagement data to form a single dataframe. Subsequently, we will merge this with the relevant products and district data, using ‘district_id’ from the district files and ‘lp_id’ from the product files. This will create a final, comprehensive dataset that will be used to analyze and visualize how engagement in learning varies across different states, tools, and digital platforms.

```
engagement <- bind_rows(engage1,engage2,engage3,engage4,engage5) # create the data frame from files in engagement data

merged <- merge(engagement, products, by = "lp_id", all.x = TRUE) # Merge the 'engagement' data frame with products data following lp_id

final_merge <- merge(merged, districts, by = "district_id", all.x = TRUE) # Merge with districts data following district_id

colSums(is.na(final_merge)) ## Check NA in each column
```

```
##                district_id                      lp_id
##                          0                          4
##                       time                pct_access
##                          0                          2
##           engagement_index                        URL
##                     135792                     256461
##               Product Name      Provider/Company Name
##                     256461                     256461
##                  Sector(s) Primary Essential Function
##                     266547                     267885
##                      state                     locale
##                          0                          0
##         pct_black/hispanic           pct_free/reduced
##                          0                          0
##   county_connections_ratio              pp_total_raw
##                     255722                     104003
```

```
knitr::opts_chunk$set(echo = TRUE)
```

c. Data cleansing & wrangling:

Once the data files are merged, the data cleansing and Wrangling is conducted to ensure the accuracy and usability of the final data frame. Once the above steps are performed, we can get to visualize and analyze the data to get the insight from the investigated data sets.

```
#CLEANING:
final_merge <- final_merge %>% drop_na(`Product Name`) # drop NA values in
column Product Name

final_merge <- final_merge %>% # Change this values into NA for cleaning
  mutate(state = replace (state, state
                          == "don\x92t know", NA))
final_merge <- final_merge %>%
  mutate(state = replace (state, state
                          == "whereabouts", NA))
final_merge <- final_merge %>%
  mutate(state = replace (state, state
                          == "NaN", NA))

## Re format percentage in column 'pct_black/hispanic':
final_merge <- final_merge %>%
  mutate (
    `pct_black/hispanic` = case_when(
      `pct_black/hispanic` == "[0, 0.2[" ~ "0 - 20%",
      `pct_black/hispanic` == "[0.2, 0.4[" ~ "20% - 40%",
      `pct_black/hispanic` == "[0.4, 0.6[" ~ "40% - 60%",
      `pct_black/hispanic` == "[0.6, 0.8[" ~ "60% - 80%",
      `pct_black/hispanic` == "[0.8, 1[" ~ "80% - 100%"
    )
  )
## Re format percentage in column 'pct_free/reduced':
final_merge <- final_merge %>%
  mutate (
    `pct_free/reduced` = case_when(
      `pct_free/reduced` == "[0, 0.2[" ~ "0 - 20%",
      `pct_free/reduced` == "[0.2, 0.4[" ~ "20% - 40%",
      `pct_free/reduced` == "[0.4, 0.6[" ~ "40% - 60%",
      `pct_free/reduced` == "[0.6, 0.8[" ~ "60% - 80%",
      `pct_free/reduced` == "[0.8, 1[" ~ "80% - 100%"
    )
  )
## Re format percentage in column 'county_connection_ratio':
final_merge <- final_merge %>%
  mutate (
    county_connections_ratio = case_when(
      county_connections_ratio == "[0.18, 1[" ~ "<1",
      county_connections_ratio == "[1, 2[" ~ ">1",
    )
  )
```

```r
## Re format percentage in column 'pp_total_raw':
final_merge <- final_merge %>%
  mutate(
    pp_total_raw = case_when(
      pp_total_raw == "[10000, 12000[" ~ "10-12",
      pp_total_raw == "[12000, 14000[" ~ "12-14",
      pp_total_raw == "[14000, 16000[" ~ "14-16",
      pp_total_raw == "[16000, 18000[" ~ "16-18",
      pp_total_raw == "[18000, 20000[" ~ "18-20",
      pp_total_raw == "[20000, 22000[" ~ "20-22",
      pp_total_raw == "[22000, 24000[" ~ "22-24",
      pp_total_raw == "[32000, 34000[" ~ "32-34",
      pp_total_raw == "[4000, 6000[" ~ "4-6",
      pp_total_raw == "[6000, 8000[" ~ "6-8",
      pp_total_raw == "[8000, 10000[" ~ "8-10",
      TRUE ~ pp_total_raw
    )
  )

## Re format values in column 'Sector(s)':
unique(final_merge$`Sector(s)`)
```

```
##  [1] "PreK-12; Higher Ed"            "PreK-12; Higher Ed; Corporate"
##  [3] "PreK-12"                       NA
##  [5] "PreK-112"                      "PPreK-12"
##  [7] "pre kindergarten to year 12"   "Corporate"
##  [9] "Higher Ed; Corporate"          "PreK-12; Higher; Corporate"
## [11] "pre kindergarten to yr 12"     "not sure"
## [13] "PreK-122"
```

```r
final_merge <- final_merge %>%
  mutate(`Sector(s)` = replace (`Sector(s)`, `Sector(s)`
                                == "PreK-122", "PreK-12"))
final_merge <- final_merge %>%
  mutate(`Sector(s)` = replace (`Sector(s)`, `Sector(s)`
                                == "PPreK-12", "PreK-12"))
final_merge <- final_merge %>%
  mutate(`Sector(s)` = replace (`Sector(s)`, `Sector(s)`
                                == "PreK-112", "PreK-12"))
final_merge <- final_merge %>%
  mutate(`Sector(s)` = replace (`Sector(s)`, `Sector(s)`
                                == "pre kindergarten to yr 12", "PreK-12"))
final_merge <- final_merge %>%
  mutate(`Sector(s)` = replace (`Sector(s)`, `Sector(s)`
                                == "pre kindergarten to year 12", "PreK-12"))


# Re format values in column 'time'/ make sure all the time in this data set
is during 2020:
final_merge <- final_merge %>%
```

```
  mutate(
    time = case_when(
      time == "1/01/2022" ~ "1/01/2020",
      time == "31/12/1020" ~ "31/12/2020",
      time == "1/1/2044" ~ "1/01/2020",
      time == "1/01/2050" ~ "1/01/2020",
      time == "1/01/2033" ~ "1/01/2020",
      time == "2/01/2050" ~ "2/01/2020",
      TRUE ~ time
    )
  )

head(final_merge) # The few lines of the data frame
```

```
##   district_id lp_id       time pct_access engagement_index
## 1        1000 12803 11/05/2020       0.00               NA
## 2        1000 10650 23/11/2020       0.00               NA
## 3        1000 32213  3/11/2020      10.40          2201.73
## 4        1000 12803  3/06/2020       0.00               NA
## 5        1000 25559 24/06/2020       0.37            47.17
## 6        1000 28525 16/09/2020       0.08             0.80
##                            URL          Product Name Provider/Company Name
## 1    http://www.slideshare.net/           SlideShare               LinkedIn
## 2           http://www.loc.gov  Library of Congress   Library of Congress
## 3  https://classroom.google.com     Google Classroom            Google LLC
## 4    http://www.slideshare.net/           SlideShare               LinkedIn
## 5 http://www.lexialearning.com/  Lexia Core5 Reading        Lexia Learning
## 6   https://www.renaissance.com Renaissance Learning  Renaissance Learning
##                       Sector(s)              Primary Essential Function
## 1            PreK-12; Higher Ed       LC - Sites, Resources & Reference
## 2 PreK-12; Higher Ed; Corporate       LC - Sites, Resources & Reference
## 3 PreK-12; Higher Ed; Corporate SDO - Learning Management Systems (LMS)
## 4            PreK-12; Higher Ed       LC - Sites, Resources & Reference
## 5                      PreK-12         LC - Digital Learning Platforms
## 6                      PreK-12         LC - Digital Learning Platforms
##         state locale pct_black/hispanic pct_free/reduced
## 1 Connecticut Suburb          60% - 80%        20% - 40%
## 2 Connecticut Suburb          60% - 80%        20% - 40%
## 3 Connecticut Suburb          60% - 80%        20% - 40%
## 4 Connecticut Suburb          60% - 80%        20% - 40%
## 5 Connecticut Suburb          60% - 80%        20% - 40%
## 6 Connecticut Suburb          60% - 80%        20% - 40%
##   county_connections_ratio pp_total_raw
## 1                       <1         <NA>
## 2                       <1         <NA>
## 3                       <1         <NA>
## 4                       <1         <NA>
## 5                       <1         <NA>
## 6                       <1         <NA>
```

```r
summary(final_merge)
```

```
##   district_id       lp_id           time             pct_access
##  Min.   :1000   Min.   :10533   Length:266995      Min.   : 0.0000
##  1st Qu.:1039   1st Qu.:30851   Class :character   1st Qu.: 0.0000
##  Median :1044   Median :55032   Mode  :character   Median : 0.0500
##  Mean   :1042   Mean   :55262                      Mean   : 0.8978
##  3rd Qu.:1052   3rd Qu.:77713                      3rd Qu.: 0.2300
##  Max.   :1131   Max.   :99916                      Max.   :77.6200
##                                                    NA's   :2
##  engagement_index       URL            Product Name       Provider/Company
Name
##  Min.   :    0.05   Length:266995      Length:266995      Length:266995
##  1st Qu.:    0.91   Class :character   Class :character   Class :character
##  Median :    5.27   Mode  :character   Mode  :character   Mode  :character
##  Mean   :  240.22
##  3rd Qu.:   37.00
##  Max.   :64818.66
##  NA's   :57546
##   Sector(s)         Primary Essential Function    state
##  Length:266995      Length:266995              Length:266995
##  Class :character   Class :character           Class :character
##  Mode  :character   Mode  :character           Mode  :character
##
##
##
##
##     locale          pct_black/hispanic pct_free/reduced
##  Length:266995      Length:266995      Length:266995
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  county_connections_ratio pp_total_raw
##  Length:266995            Length:266995
##  Class :character         Class :character
##  Mode  :character         Mode  :character
##
##
##
##
```

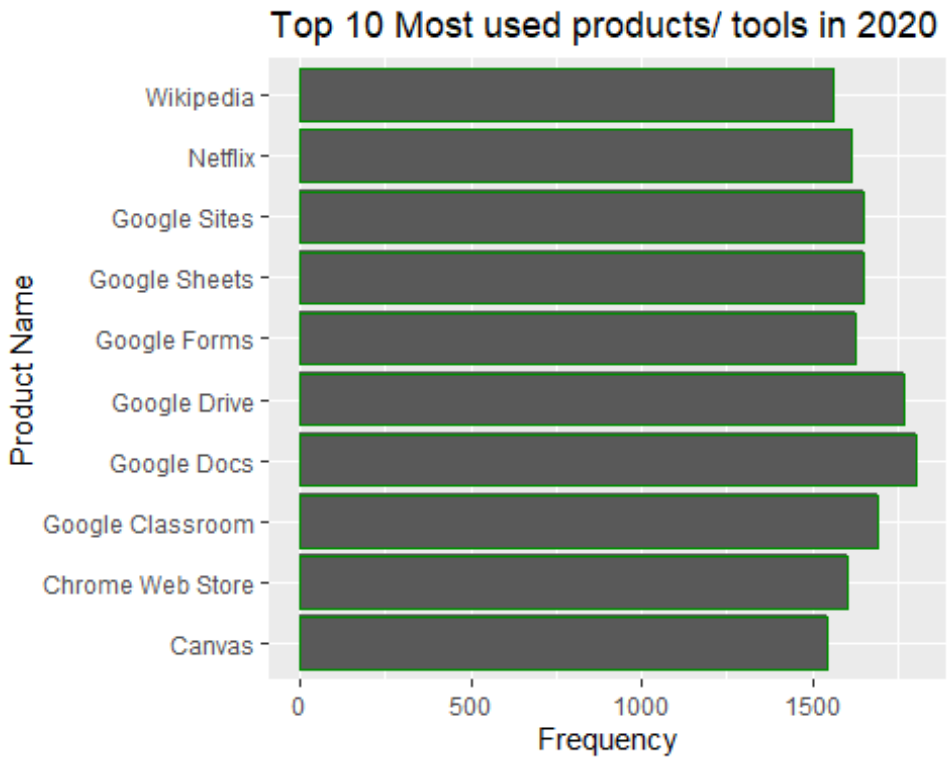
```r
knitr::opts_chunk$set(echo = TRUE)
```

2. Data Visualisation:

2.1. Products/ Tools most used across states:

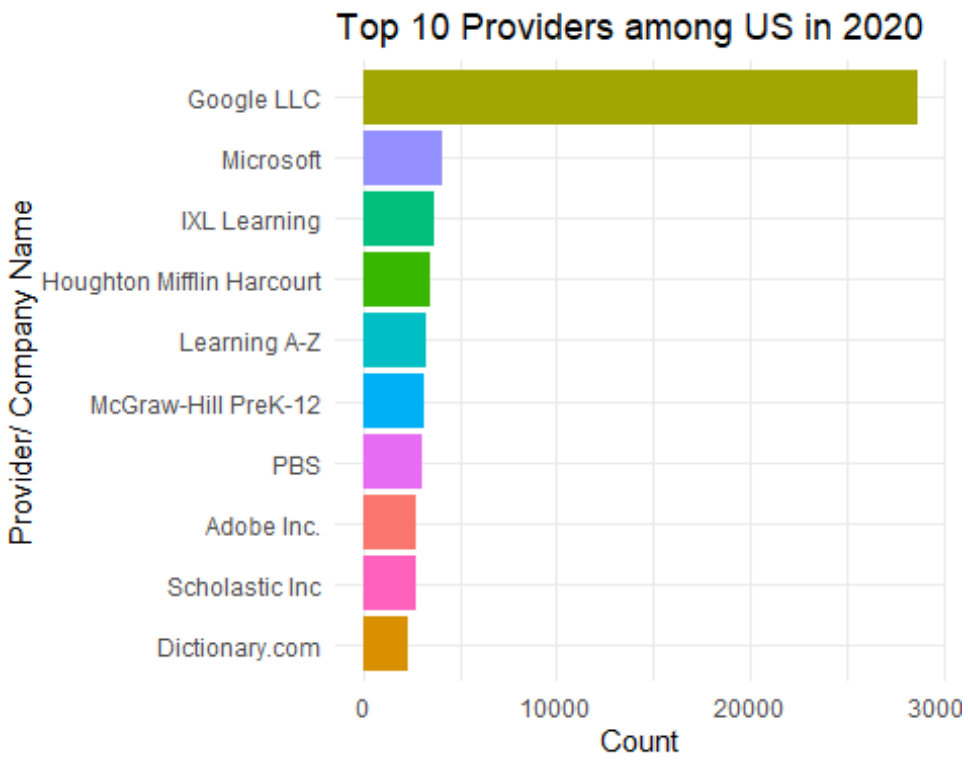
a. Most used products among students across states:

To illustrate the most used products in 2020, a bar graph is particularly suitable as it offers a clear and direct comparison of the data. In this dataset, the products or tools frequently used in digital learning will each be represented by individual bars, providing a straightforward comparison of the most popular tools.



b. Top providers in digital learning across US:

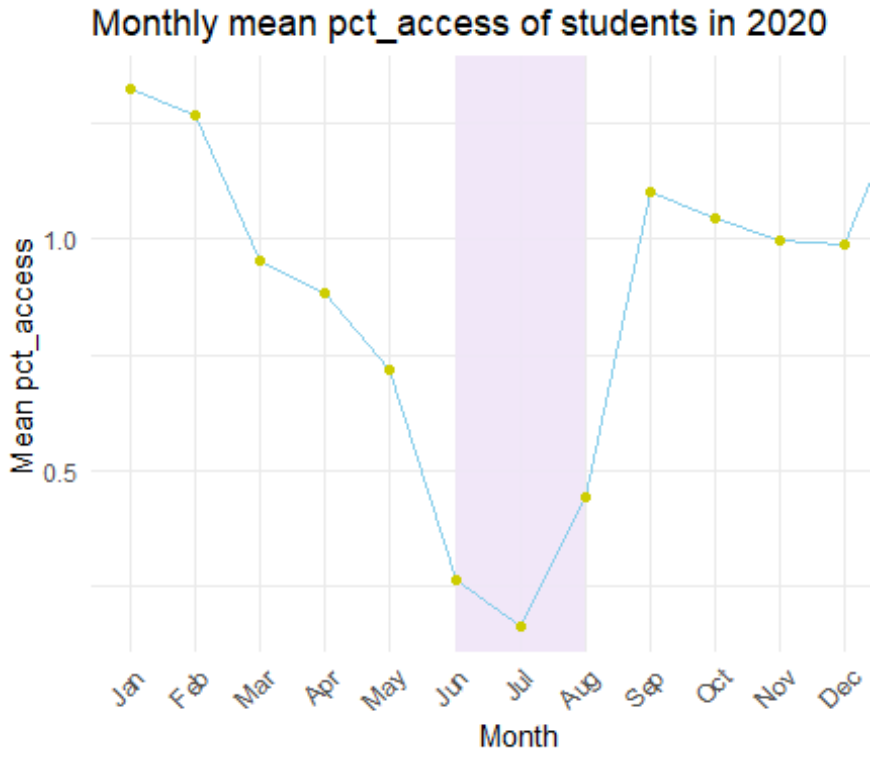
The bar graph chosen here effectively highlights the leading company in supplying digital tools and platforms for educational purposes across five districts in the U.S. It aims to showcase the comparative usage of digital tools provided by different companies for online learning. Additionally, the graph reveals which company offers the highest number of products that are widely adopted by students in these districts.



2.2. Student engagement and digital learning access:

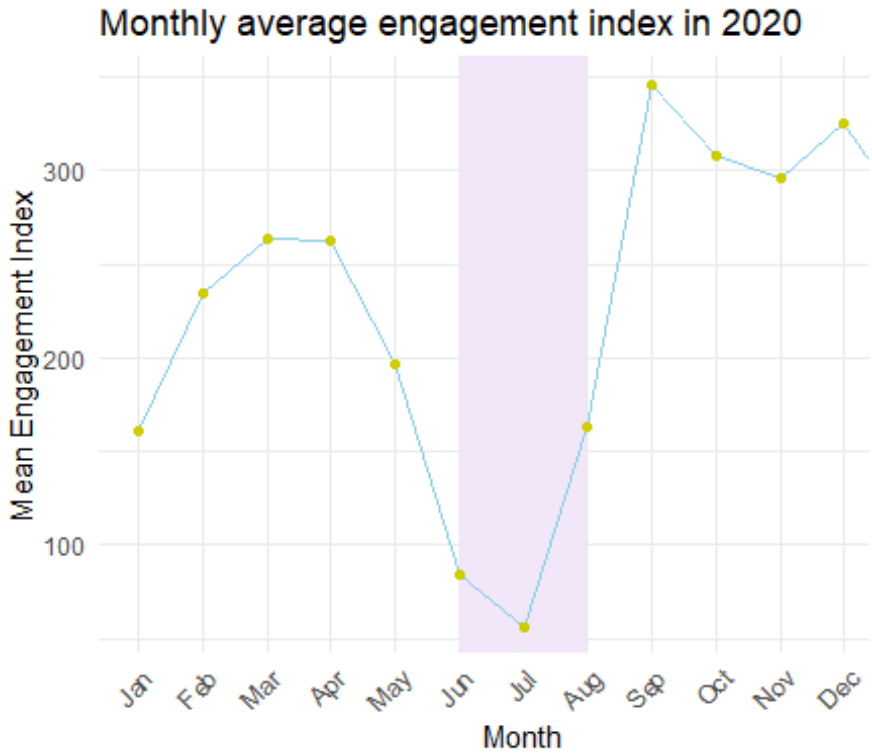
a. Monthly percentage of students accessing online learning tools:

The following visualization represents the average mean percentage of students accessing digital learning tools. Then, a line graph tends to be the most optimal graph excellently displays trends over time. This type of graph presents a clear visual tracking of student assessment with digital tools throughout a specific time (during 2020). Besides, line graphs excel in showing increases, decreases, and plateaus, making it efficient to identify patterns or significant changes at a glance. This is crucial for companies providing digital platforms for online learning to understand temporal variations in student engagement, helping them to pinpoint times of higher or lower usage and adjust strategies or resources accordingly.



b. Monthly engagement of students with digital learning tools:

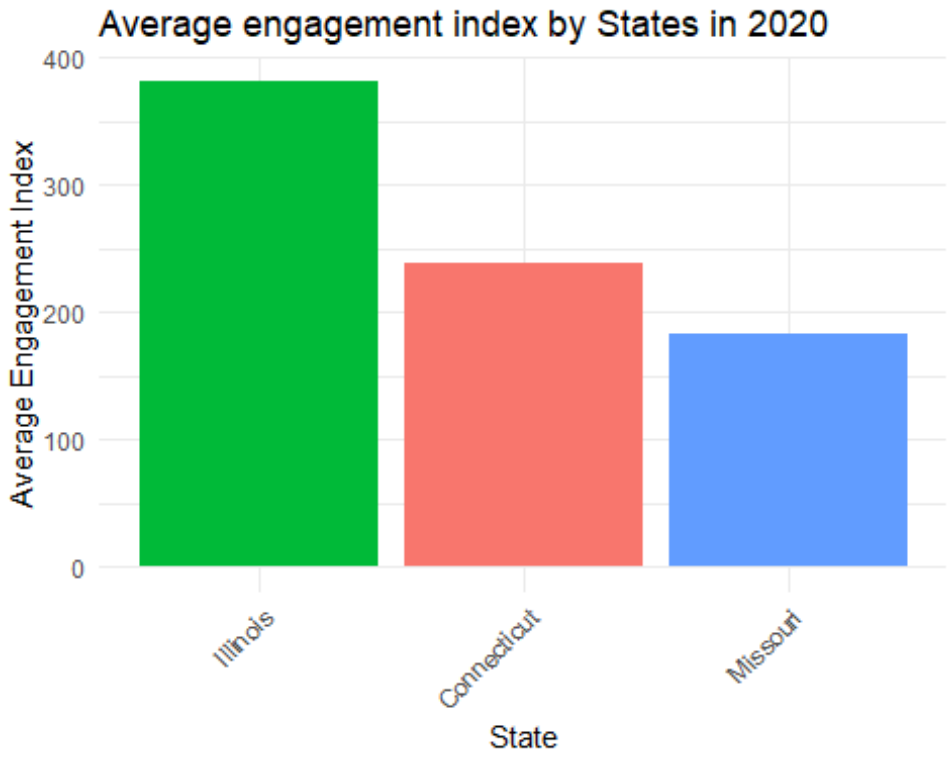
The visualization below displays the average engagement index of students with digital learning tools across districts using a line graph, similar to the earlier student access percentage. This format effectively shows trends and changes over time.



2.3. Engagement of students in digital learning in different states:

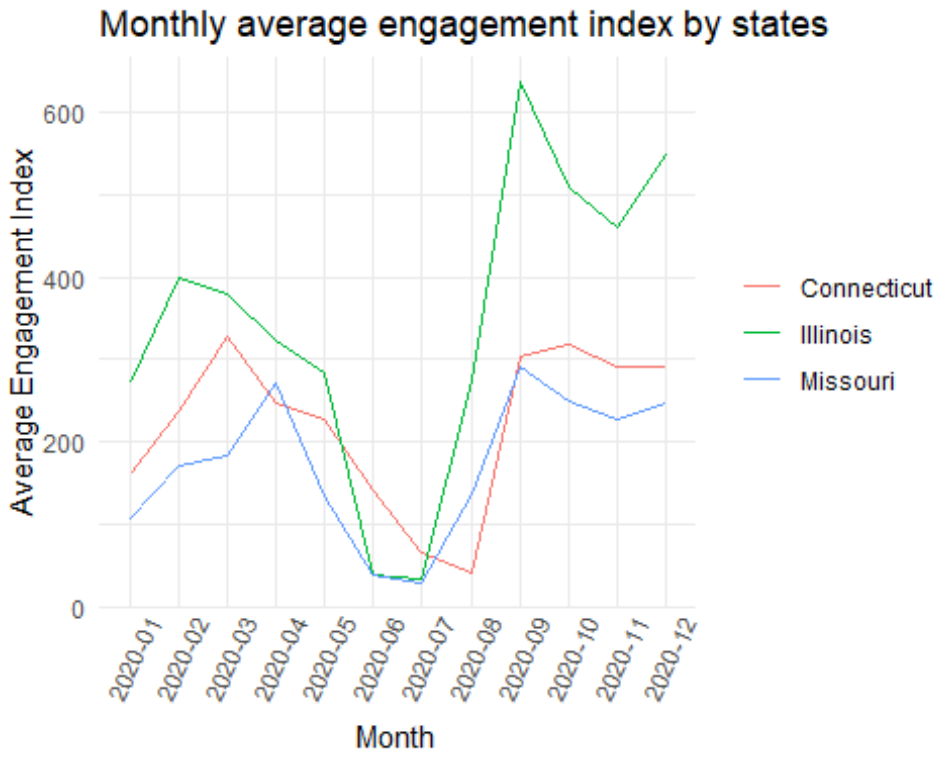
a. Engagement index by states:

This visualization employs a bar graph to compare the average engagement index of students, based on at least one daily product page-load event, across different states. The bar graph is particularly effective for this purpose as it clearly delineates differences between states, allowing for direct and straightforward comparisons. However, it should be noted that districts 1039 and 1131 are excluded from this analysis due to the missing information of state name and other relevant data in the district files.



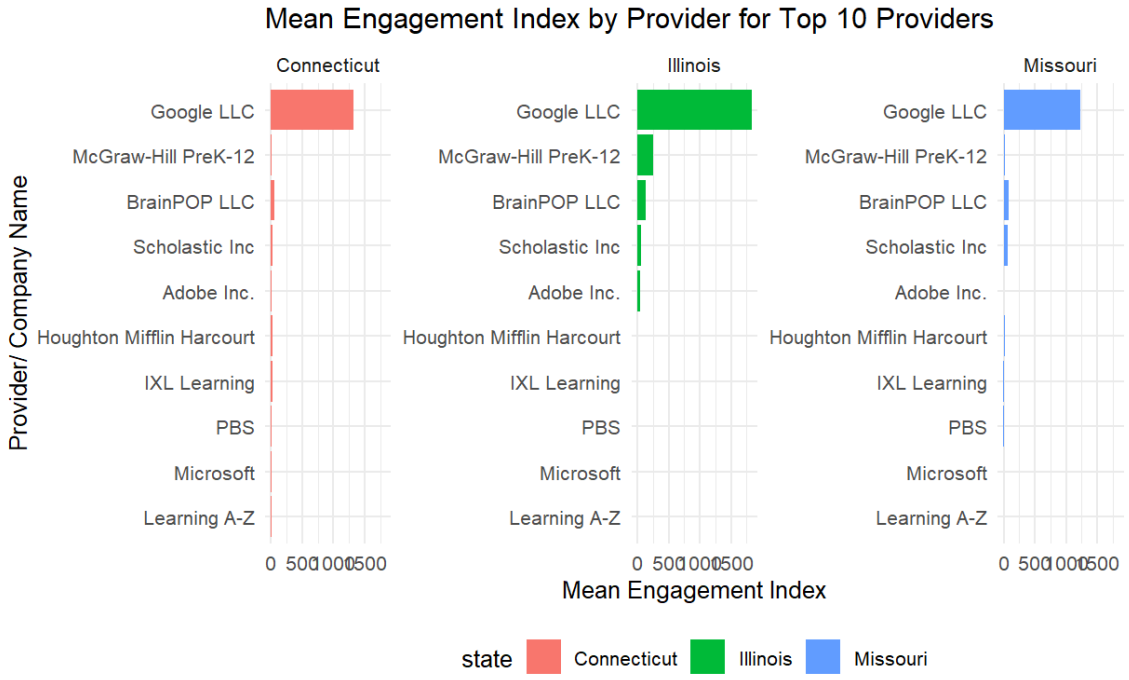
b. Engagement index by state monthly:

This following visualization follows a line graph format. This graph effectively captures the trend in student engagement indices across various states throughout 2020. The use of a line graph is particularly advantageous for this analysis as it highlights temporal changes and facilitates the identification of patterns over time, making it ideal for observing trends in data across a specified period.



2.4 Engement index with provider in digital learning by states:

To investigate the relationship between student engagement and digital tool providers, we will use a series of three bar graphs. These graphs will clearly compare student engagement across the top 10 providers in three states, effectively highlighting differences.



3. Findings:

3.1. Products/ Tools most used across states:

As the COVID-19 pandemic forced schools to shift from face-to-face to online teaching, a significant rise in the use of digital tools occurred within the educational sector. During this time, a variety of digital platforms became essential for adapting to the new demands of remote education. The analysis of the top 10 products and providers clearly shows Google's dominance in the online learning landscape. According to the data, Google not only features prominently but also accounts for seven out of the top ten most utilized products across schools in the U.S. This significant presence underscores its leading role in equipping students with necessary digital resources.

However, it's worth noting that Google's strong showing in the top used online tools is not entirely surprising. Given its pre-pandemic prevalence in both personal and professional spheres, its continued prominence in the educational sector during the crisis was expected, reflecting its already established base rather than a sudden surge in utility or popularity.

3.2. Student engagement and digital learning access:

The analysis of digital learning trends in 2020 showcases a marked seasonality in student engagement with educational platforms. The engagement index and the percentage of students accessing digital tools followed a similar downtrend during the summer months (practically from June to August), indicating a decrease in digital learning activities during the school break. However, a notable difference emerges in the timing of peak activities; while the highest percentage of students accessing the tools occurred in January, the engagement index reached its peak later in the year, in August. This disparity highlights a divergence in how frequently students log onto digital platforms versus how intensively they engage with content, suggesting that while fewer students may initiate sessions at certain times, those who do are more actively involved.

3.3. Engagement of students in digital learning in different states:

To analyze student engagement across different regions in the US, this study compares the engagement index among students in Illinois, Connecticut, and Missouri. The data reveal that students in Illinois engage more intensively with digital learning content, as evidenced by an engagement index significantly higher than that observed in the other two states. Additionally, the trend in the engagement index across these states mirrors the overall pattern observed in the broader dataset of investigated districts, with a notable decline during the summer break. This similarity suggests that student engagement consistently decreases during this period across various states.

3.4. Engement index with provider in digital learning by states:

Google emerges as the dominant digital platform provider across Illinois, Connecticut, and Missouri, consistently leading in engagement. Specifically, in Illinois, Google not only leads but does so by a significant margin, with an average daily page load event nearing 1500,

underscoring its popularity. Although McGraw-Hill ranks far behind Google, it holds the position of the second most utilized provider in Illinois.

In conclusion, Google's dominance in the digital learning space is evident across the studied states, with Illinois showcasing particularly high engagement levels. Despite a considerable gap, McGraw-Hill remains a significant secondary provider within Illinois.