

Team 97

Katrina Gutierrez, Isaac Kerson, Courtney MacLellan, Richie Phan

## Visualization

**Demographics determined the likelihood that a student had contact with a teacher during the pandemic.**

**Student Traits**

Race: Hispanic

Income: 50-75k

Internet availability: Sometimes

Computer availability: Sometimes

Submit

Visit <http://13.91.125.11:8501/> to view an interactive version of the dashboard.

**47.1%**



Likelihood that **Hispanics** who **sometimes** have internet, **sometimes** have a computer, and live in a household with an income of **50-75k** had virtual contact with a teacher **four or more times** per week.

## Motivation

### Problem Statement

Our goal is to explore and visualize the digital divide, specifically the impact of different combinations of factors on the amount of live virtual contact students had with teachers during the COVID Pandemic using the U.S. Census Household Pulse Survey data.

### Why should we care?

The effect of the digital divide on education is important because of its potential long-term effect on economics and the quality of life for future generations. Digital resources are essential for students facing school closures during the pandemic to continue their education.

## Approach

### What are your approaches?

We built an ordinal logistic regression model that could be used to predict and visualize the number of days of live virtual contact students had with teachers during the 2020-2021 school year. Our final model was built in R using the MASS library and included the predictors Race/Ethnicity, Income, Internet Availability, and Computer Availability.

The visualization was created using Streamlit in Python and uses the model output to demonstrate in an instantly coherent way how the selected independent variables correlate with the probability that segments of the population receive four days of virtual live instruction per week.

### How does your approach work?

Users can adjust key demographic features and see how they change the likelihood that a student received an adequate education during the pandemic. Lay users as well as experts will gain a deeper, intuitive understanding of the digital divide by experimenting with the dashboard.

### Why do you think your approach can solve the problem? (Intuition)

Our tool provides policy makers and educators a more complete understanding of the factors that influenced students' access to digital education during the COVID pandemic beyond obvious explanations like internet and computer availability. The visualization allows users to explore the factors affecting the disparities in virtual education and gain a deeper understanding of students' learning experiences during the 2020-2021 school year.

### Innovations:

- Used 6 months more data than previous research.
- Investigated live virtual contact with teachers while most visualizations focus on access to computers / internet.
- Visualization allows users to investigate the effect of different combinations of variables, rather than a single factor at a time.

## Data

### How did you get the data?

Consists of biweekly online survey conducted by the U.S. Census Bureau. CSV files were downloaded for weeks 13-33 (August 19 2020 to June 23 2021).

Data source: [U.S. Census House Pulse Survey Website](#)

### Data Characteristics

We chose this range to correspond with full school year and merged all the files into 3 gigabyte file. According to the U.S. Census website the survey was an online 20-minute suvey and people were randomly chosen to participate across the country. The collection period was split into three phases.

## Experiment and Results

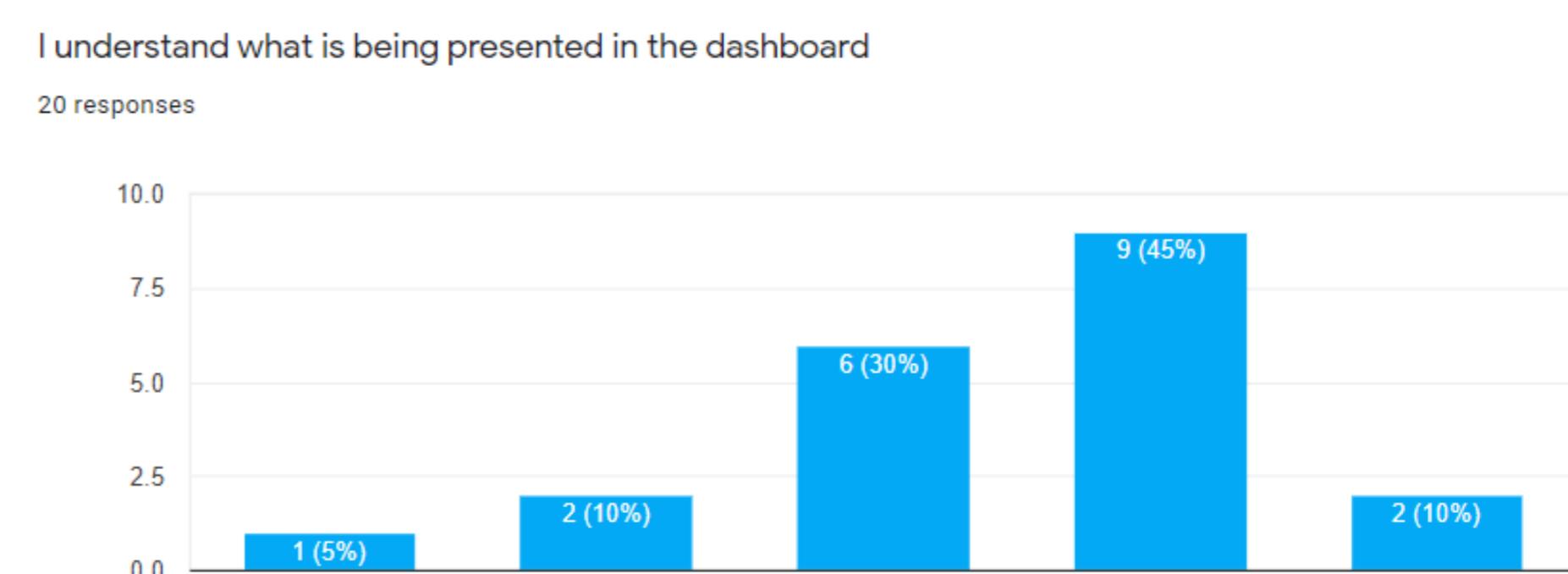
### Survey

#### How did you evaluate your approach?

To evaluate and improve upon the first version of our visualization, we conducted a user survey. We wanted to determine ease of use, interpretability, and if we needed to add more predictors.

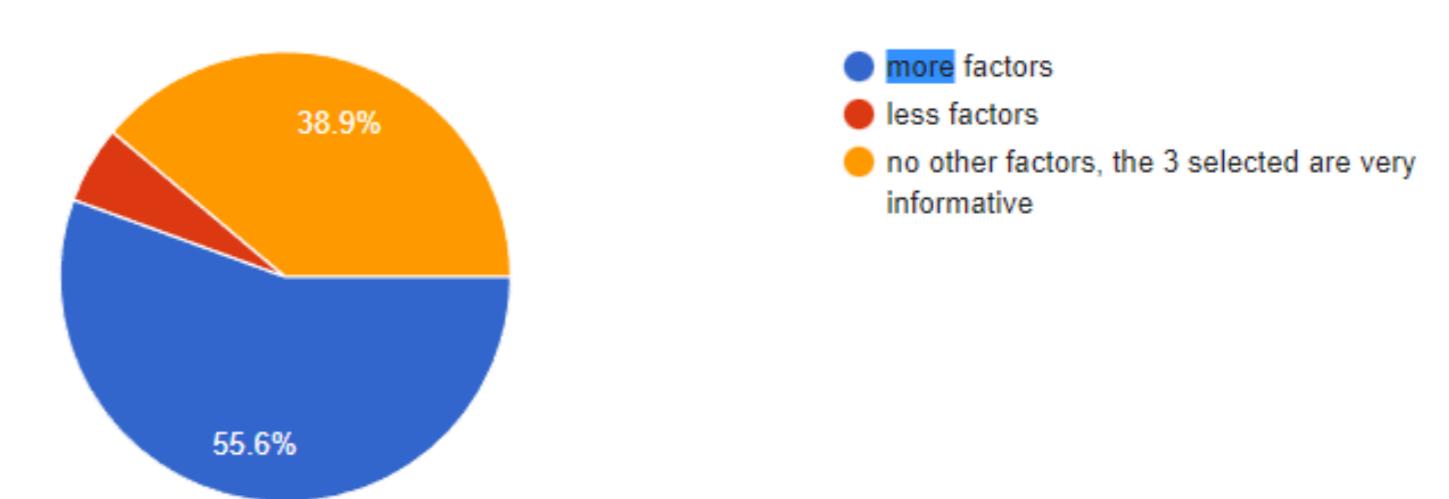
#### What were the results?

Based on the feedback from 20 responses, slightly more than half (55%) agreed that they understood what the visualization presented. More than half of respondents preferred more factors than the initial 3 represented.



Aside from race, income, and internet, I would have preferred if the dashboard explored...

18 responses



We also received and incorporated the feedback below:

- Quartile comparison is confusing and should be removed.
- There should be instructions on how to navigate the dashboard.
- Smoothen transition when altering factors.

### Model Evaluation

#### How did you evaluate the model?

To evaluate our model and choose the best predictors, we compared it with other ordinal logistic regression models with other predictors using accuracy, likelihood ratio tests, and AIC. In the first phase of our project, we only used 3 predictors (race/ethnicity, income, internet availability) and we wanted to test out adding one more due to our user survey results having a preference for more predictors (55.6% vs 38.9% preferring the current 3).

We compared 5 models in total, the original model (M1) with three variables and four others with the original 3 predictors with one additional (computer availability (M2), food sufficiency (M3), education level of respondent (M4), or total number of kids in households (M5)).

#### What were the results and how do your methods compare to others?

The models we compared all had an accuracy around ~60%. The models with 4 predictors were all significant improvements upon a model with just 3 predictors based on the likelihood ratio tests (which all came out significant). For AIC, our final (chosen) model had the lowest score of 455403.7.