Correlation between School Choice and COVID-19 Positive Case Rates

Final Project
CS620 Intro to Data Science and Analytics
John Romano
30 November 2020

Background and Objective

OLD DOMINION UNIVERSITY
IDEA FUSION

- Due to ongoing Corona Virus Disease 2019 (COVID-19) pandemic, school districts in Virginia (VA) faced with decision of how to proceed with K-12 education for Fall 2020
- According to VA Dept. of Education (VDE), Municipalities chose between three schedule categories:
  - Fully/Partially In-person
  - Fully/Partially Hybrid (blended in-person and remote)
  - Fully Remote
- **How do these decisions compare with trends in COVID-19 infection history in VA?**

Process

OLD DOMINION UNIVERSITY
IDEA FUSION

- Determine datasets
- Pre-process datasets
- Load and clean datasets
- Combine pertinent data
- Visualize data and perform analysis
- Make conclusions and determine path forward

Datasets

OLD DOMINION UNIVERSITY
IDEA FUSION

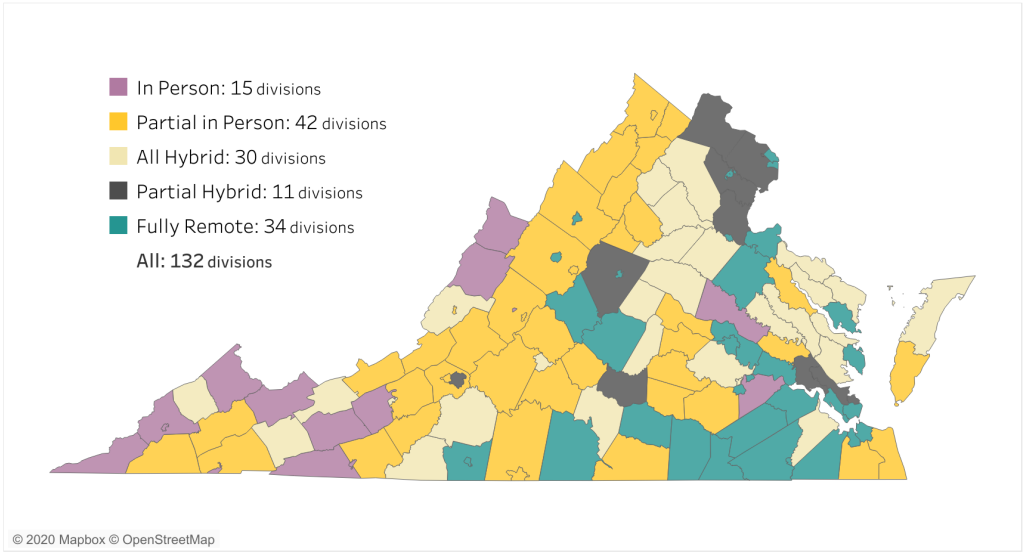
- VDE dataset with information on 2019 enrollment, school choice, and learning schedule/number of days in person
  - Data current as of 9/22/2020
- VA Dept. of Health (VDH) datasets containing total PCR testing events and positive COVID-19 cases to date
  - Updated daily
  - Presented in separate datasets in terms of Zip Code, Age Group and Health District
  - Data current as of 11/9/2020
- Census data with population projections for 2019
- Zip Code-to-Municipality Name mapping
- Municipalities contained within each VA Health District

Goal of building one DataFrame with pertinent VDE and VDH information

Dataset Pre-processing

- VDE data is available online, but not available for download
  - https://www.doe.virginia.gov/support/health_medical/office/reopen-status.shtml
  - Pertinent data from above manually replicated into an Excel spreadsheet to allow for inclusion in this study
- Composition of VA Health Districts available as PDF from VDH website
  - PDF parsing proofed to be a difficult task, so data manually replicated into Excel for expedience
- Remaining data available as CSV files and required no pre-processing

Virginia's Return to School Instructional Schedules Fall 2020



In Person = 4+ days of in person instruction for all students
Partial in Person = 4+ days in person for some students (usually the younger grades); hybrid or remote for all other students
All Hybrid = all students with some in person and some remote (none hitting the 4 days/week threshold)
Partial Hybrid = some students hybrid (usually the younger grades; none hitting the 4 days/week threshold), all others fully remote
Fully Remote = learning is remote for the vast majority of students, while some students may have in-person learning experiences

Hybrid schedule description
A = Group A
B = Group B
C = Group C
D = Group D
R = Remote

NOTE: All Virginia school divisions have offers students and families a fully-remote option to learn.

Importing and Cleaning Datasets

OLD DOMINION UNIVERSITY
IDEA FUSION

- All datasets imported as Pandas DataFrames using either CSV or Excel data readers
- Datasets are processed to replace any missing information with NaN or 0, depending on the application
  - For example, missing PCR testing data would be replaced with NaN while missing info on number of days in school would be replaced with 0
- Index labels updated as required for additional clarity and ease interfacing between different datasets

Combining Datasets

OLD DOMINION UNIVERSITY
IDEA FUSION

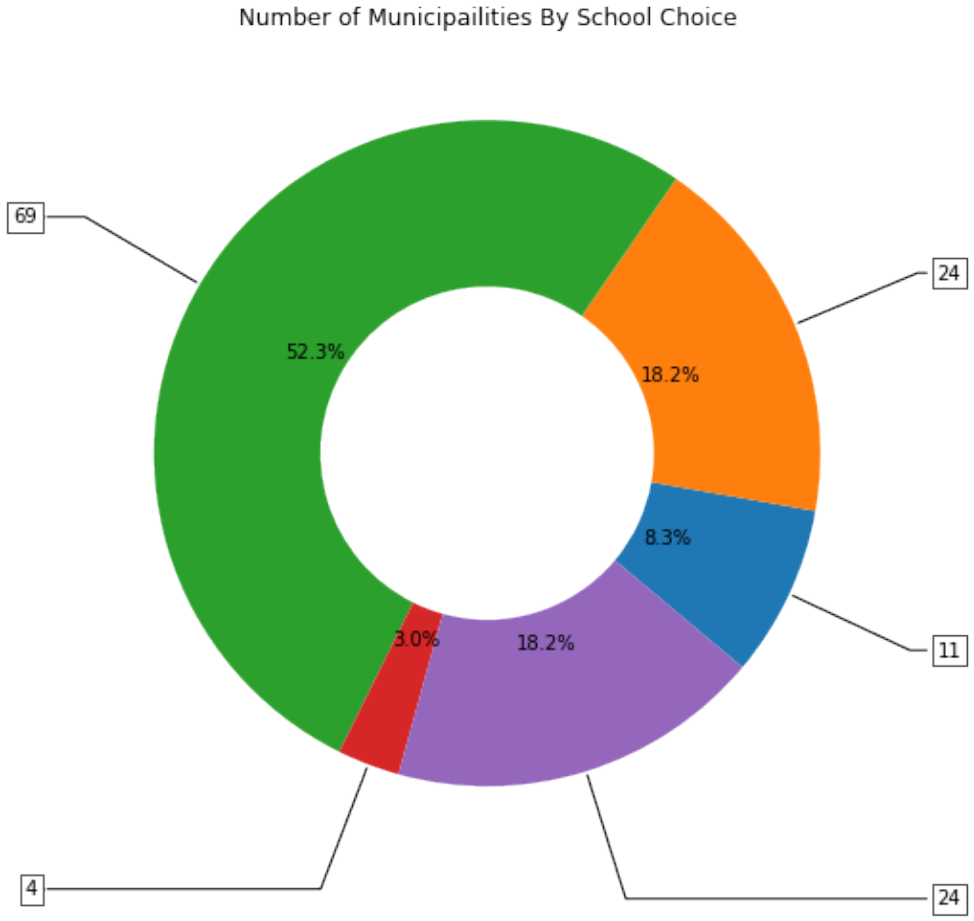
- Several functions written to parse pertinent information from VDH and Census data for inclusion in VDE datasets
  - Number of positive cases and testing encounters for a particular date
  - 2019 population estimate for Census data
  - Functions to map zip code, health district and town name to City/County format used in VDE dataset
- Add new columns to VDE dataset for Testing and Positive Cases for 8/1, 9/1, 10/1 and 11/1 to assess trends in positive case rate
  - Population data proved not useful due to conversion issues from Town to County name
- Data also collected for above dates across Health District by Age Group in separate dataset to assess trends in positive case rate

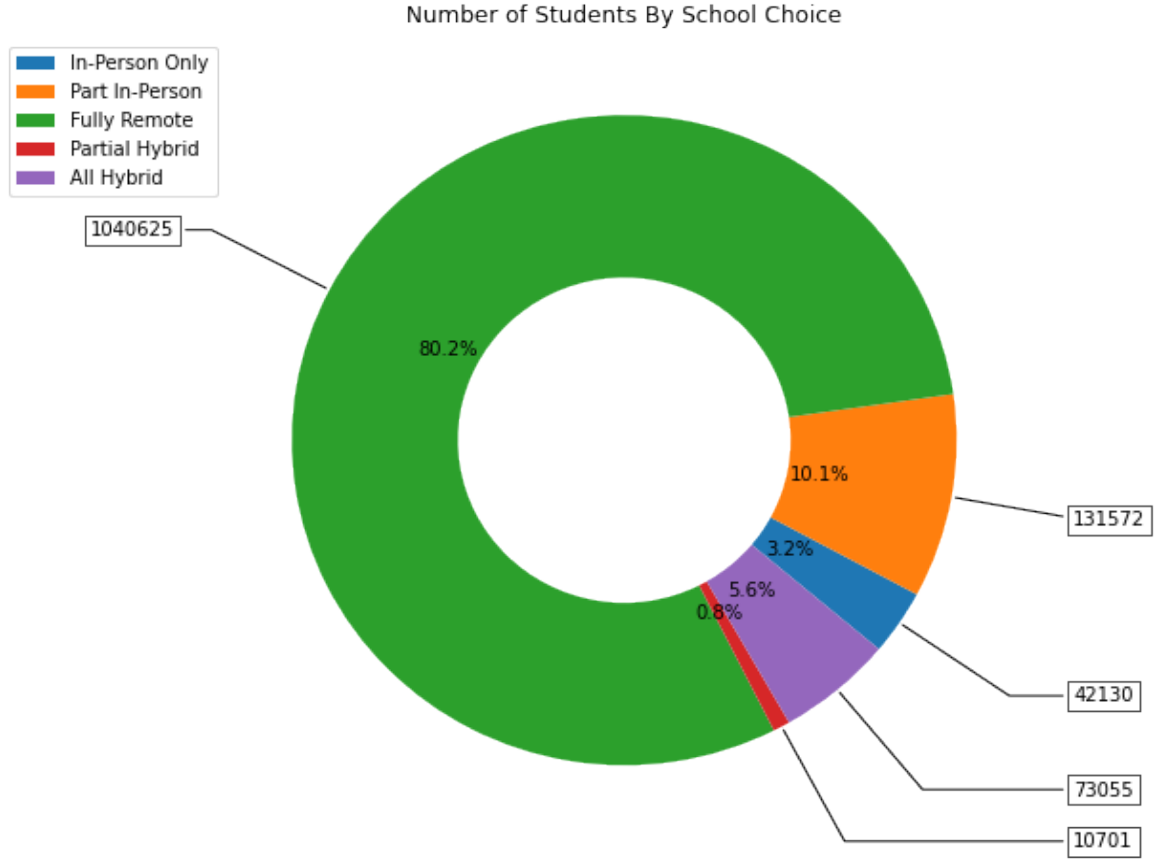
Data Visualization

OLD DOMINION UNIVERSITY
IDEA FUSION

- Several visuals created to learn trends within the data
  - Distribution of school choice by municipality and by enrollment
  - Number of days in Person by grade level
  - Positivity rate in Age Group 0-19 (school aged) by Health District

Data Visualization – School Choice Distribution

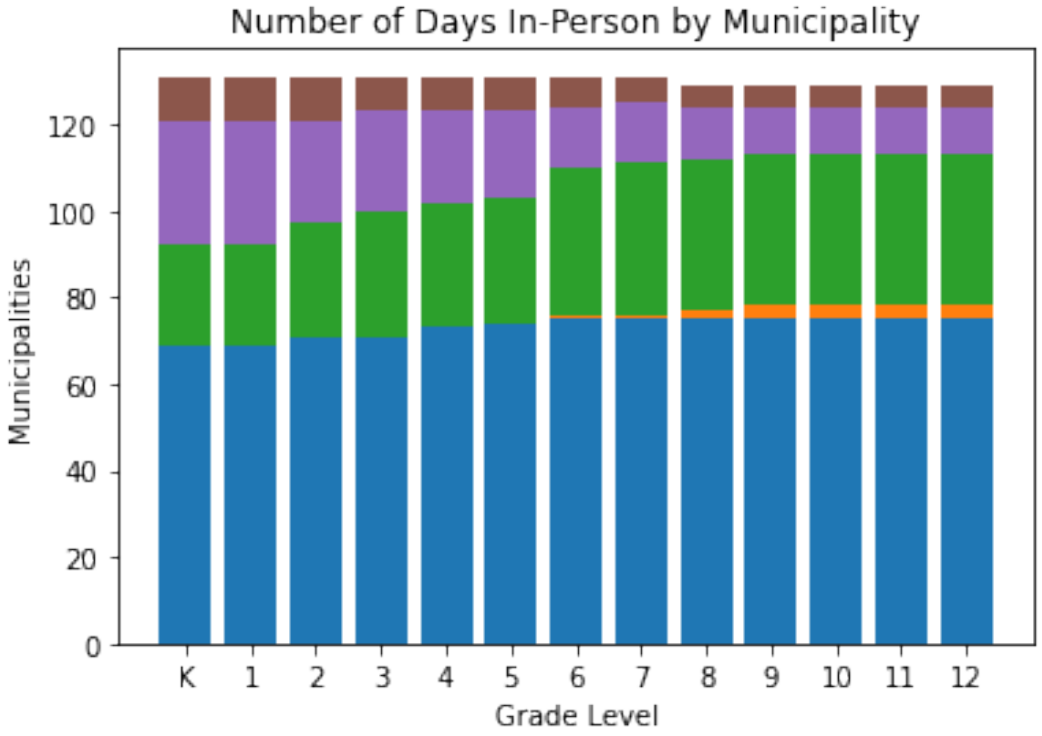


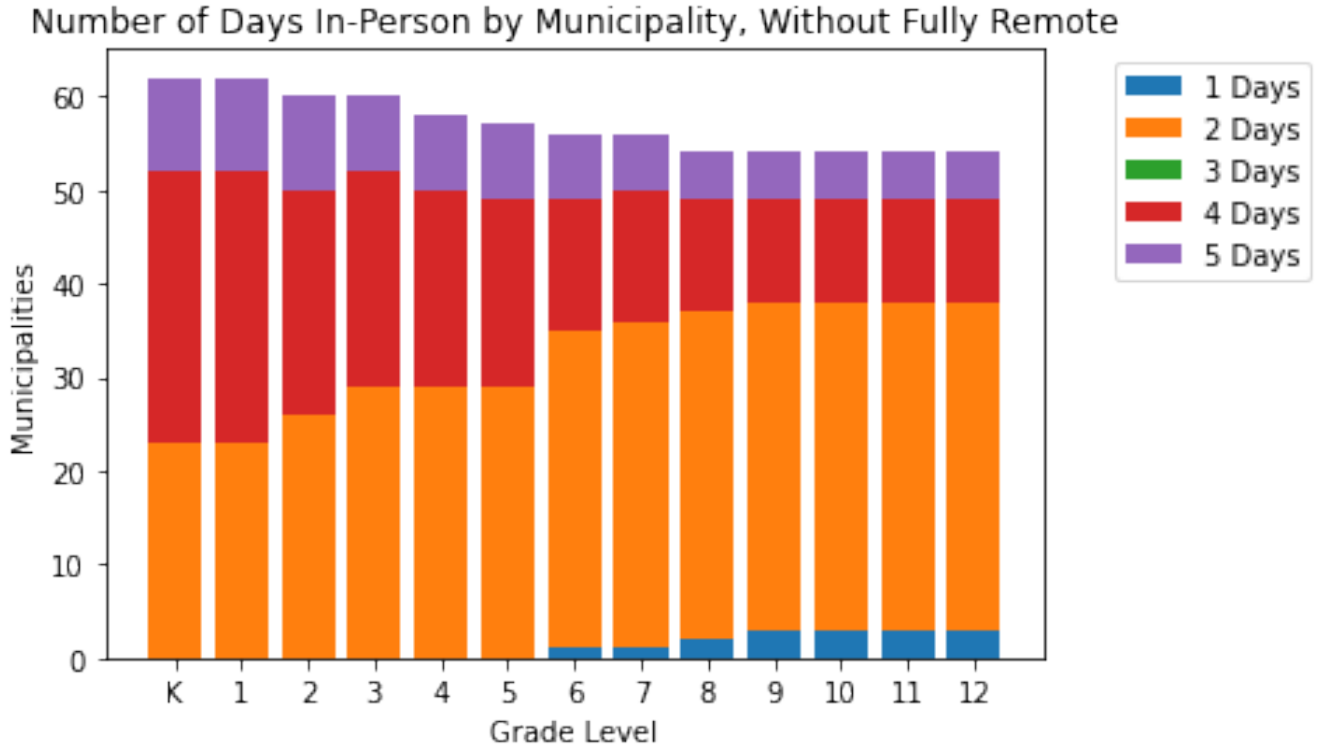


Majority of students enrolled in fully remote

Municipalities with higher enrollments tend to be fully remote

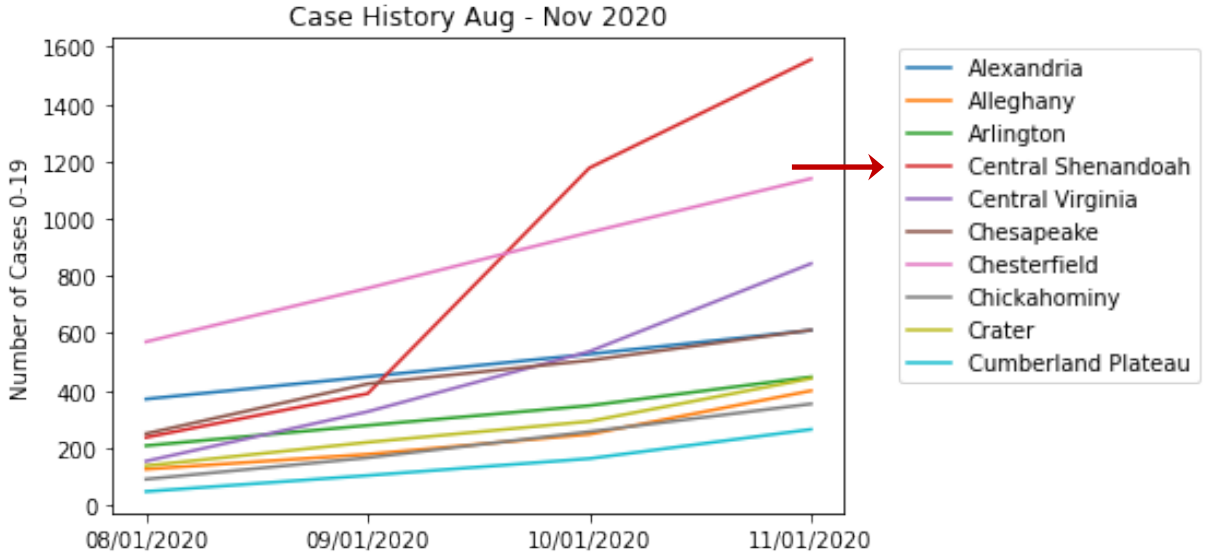
Data Visualization – Distribution Number Days In-Person

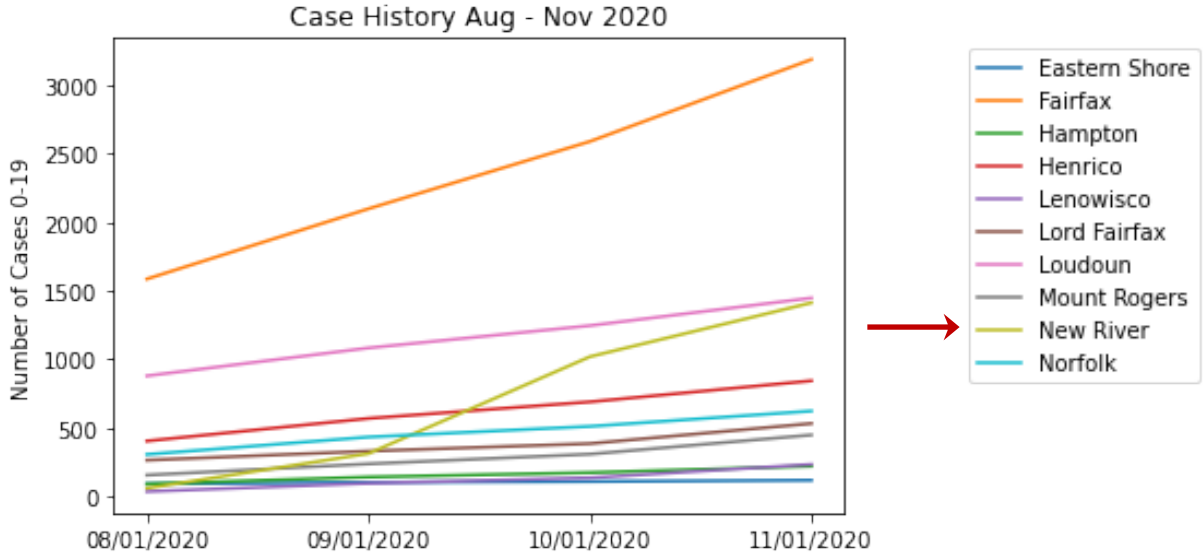


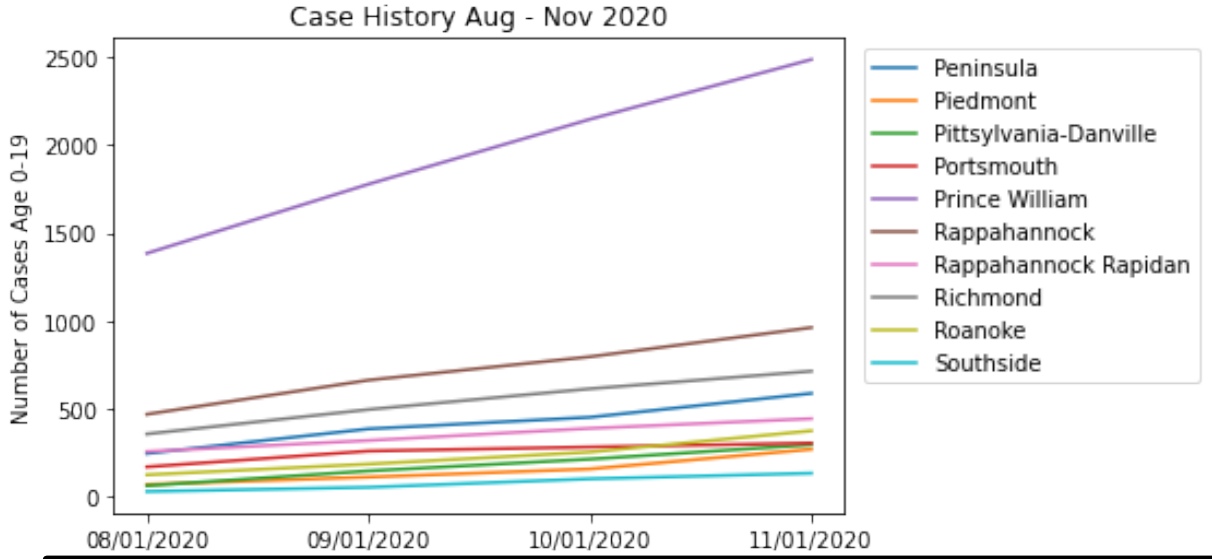


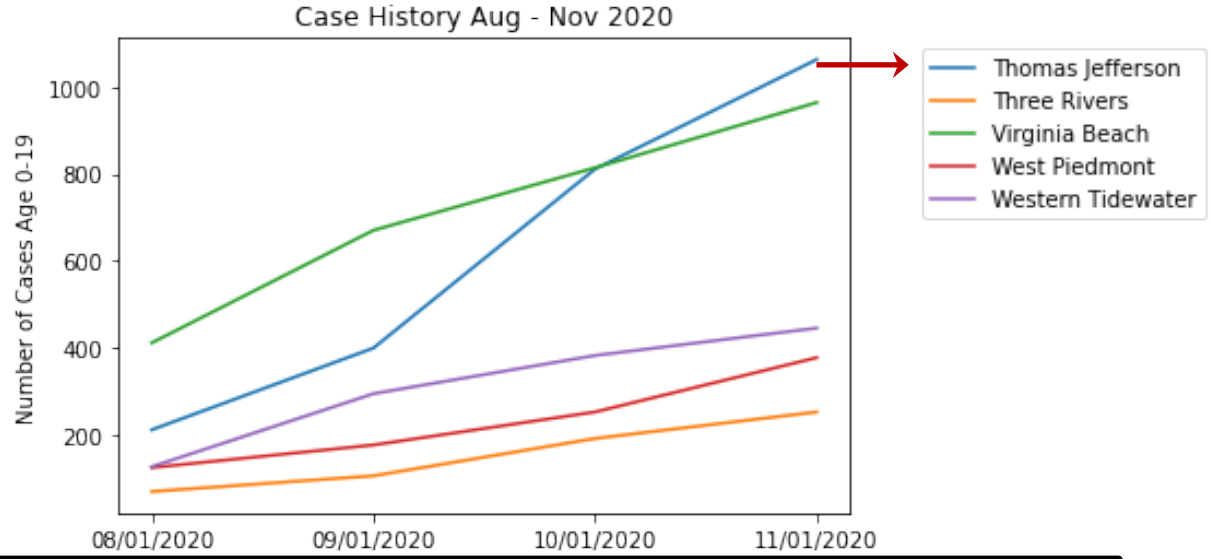
Lower grades tend to be in-person more

Data Visualization – School Choice Distribution





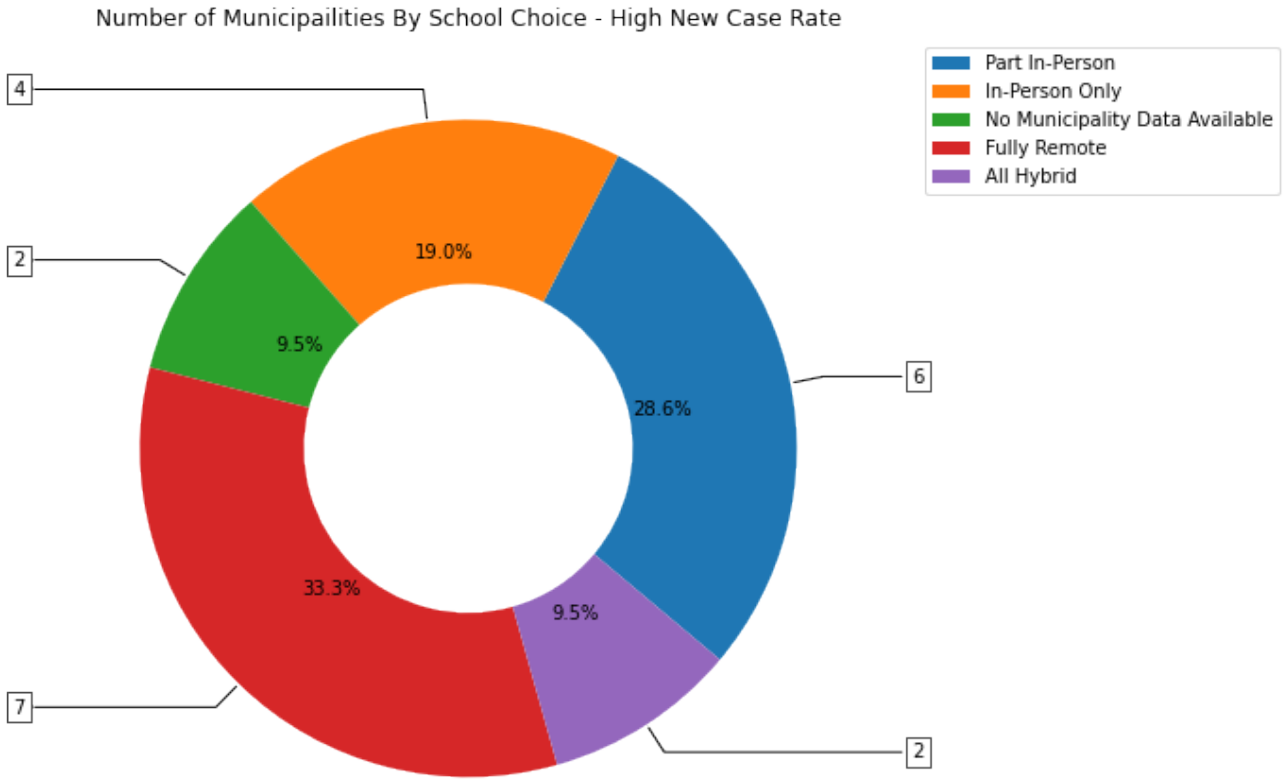




Three Districts with increased positive rate from September to October

Data Analysis

- Give more attention to districts with increased positive case rate
  - Consider school choice – 36.8% Fully Remote in high positive rate vs 52.3% in general population



Note: two municipalities not properly plotted due to NaN

	CityCounty	Choice
8	Augusta	Part In-Person
9	Bath	In-Person Only
17	Buena Vista City	NaN
57	Harrisonburg	Fully Remote
60	Highland	In-Person Only
69	Lexington	In-Person Only
108	Rockbridge	Part In-Person
109	Rockingham	Part In-Person
118	Staunton	Fully Remote
126	Waynesboro	Fully Remote
41	Floyd	Part In-Person
48	Giles	All Hybrid
81	Montgomery	Part In-Person
101	Pulaski	In-Person Only
102	Radford	All Hybrid
1	Albemarle	Fully Remote
23	Charlottesville	Fully Remote
42	Fluvanna	Fully Remote
52	Greene	Part In-Person
71	Lousia	NaN
82	Nelson	Fully Remote

Majority of Municipalities with increased positive rate had some in-person component

Conclusions

OLD DOMINION UNIVERSITY
IDEA FUSION

- Project provided opportunity to leverage skills learned throughout course to perform a real-world analysis
- Gained experience manipulating and combining various datasets from different sources
- Gained experience with Python toolkits such as Pandas, Numpy and Matplotlib
- Tracked general COVID-19 testing and positive case trends throughout VA from August to November 2020
  - Able to show several Health Districts which show an increased positive case rate for school-aged persons coinciding with schools starting for the Fall
  - Districts with increased positive case rate exhibit lower percentage of fully remote school than the general population

Correlation likely exists between in-person school and increased positive rate in Health Districts studied

Path Forward

- Look at more current data
  - VDE data available for Nov 12
  - VDH data updated daily
- Consider updated trends as Municipalities change their learning systems throughout the year
- Consider other causes that may affect positive case rate
  - Holiday travel, weather conditions, public health posture/reopening phase

Questions?

- Thank you for your attention!
- Any questions about this presentation can be emailed at jroma013@odu.edu