Talk

COVID-19

PERSPECTIVES FROM A DATA SCIENTIST

Talk by **Dr. Rick Lawrence**, a Data Scientist who currently serves on the Ridgefield COVID-19 Task Force. Dr. Lawrence's work has evolved from nuclear science to computer science to machine learning and, most recently, to quantitative finance. He joined IBM Research in Yorktown Heights, NY in 1987, where he held a number of management positions, most recently as Distinguished Research Staff Member and Senior Manager, Machine Learning & Decision Analytics. From 2016 to 2019, he was President of PCIX, Inc, a New York City VC-funded startup that used machine learning to extract quantitative insight on the relationship between private-equity transactions and the performance of public markets.

November 3 5:30 - 7 p.m.

Science Building 125, Midtown campus

Free and open to the public

For more information, call Dr. Anna Malavisi at (203) 837-3271.

WESTERN CONNECTICUT STATE UNIVERSITY

We will address the following key questions from a pure data-science perspective:

1. What is the current state of the pandemic in the U.S. and Connecticut?
2. Based on the latest vaccination data, will the U.S. reach herd immunity?
3. Do masks reduce COVID infection rates?
4. What is the relationship between vaccination rates, natural immunity, and new case rates?

SCAN HERE TO REGISTER

COVID-19: Perspectives from a Data Scientist

Rick Lawrence, Ph.D.

Data Scientist
Ridgefield COVID-19 Task Force

Western Connecticut State University
November 3, 2021

COVID-19: Perspectives from a Data Scientist

I. COVID-19 Data Science: a personal journey

II. Introduction to COVID-19 data reporting & data science framework

III. We address the following questions from a pure data-science perspective:
1. What is the current state of the pandemic in the US and Connecticut?
2. Based on the latest vaccination data, will the US reach herd immunity?
3. Do masks reduce COVID infection rates?
4. What is the relationship between vaccination rates, natural immunity, and new case rates?

IV. Personal Perspectives: On the convergence of COVID Data Science and Political Science

So how did you become a COVID Data Scientist?

Formal Education

B.S. Chemical Engineering — Stanford University

Ph.D. Nuclear Engineering — University of Illinois

Nuclear Science (1979-1987)

Computational Neutron Transport Methods — Argonne National Laboratory

Borehole Nuclear Scattering Methods — Schlumberger-Doll Research

Computational Science → Machine Learning (1987 – 2016)

High-Performance Computing → Data Mining → Predictive Modeling → Machine Learning & Decision Analytics

IBM Research

Quantitative Finance (2016-2019)

Private Equity Transitions → Insight on Public Markets

PCIX Inc

Covid Data Science (2020 – ????)

COVID Data Science

Town of Ridgefield

Previous Talk at *Western Connecticut State University*:
"*On the Convergence of Machine Learning and Artificial Intelligence*"
March 29, 2017

COVID-19: Perspectives from a Data Scientist

I. COVID-19 Data Science: a personal journey

II. Introduction to COVID-19 data reporting & data science framework

III. We address the following questions from a pure data-science perspective:

1. What is the current state of the pandemic in the US and Connecticut?
2. Based on the latest vaccination data, will the US reach herd immunity?
3. Do masks reduce COVID infection rates?
4. What is the relationship between vaccination rates, natural immunity, and new case rates?

IV. Personal Perspectives: On the convergence of COVID Data Science and Political Science

We start with a Basic Primer on COVID data reporting

- The ***New Case Rate*** is the number of new cases per day per 100,000 Residents
  - It is typically averaged over the past 7 days to remove weekend fluctuations.
- The ***New Test Rate*** is the number of new tests per day per 100,000 Residents
  - It is typically averaged over the past 7 days to remove weekend fluctuations.
- The ***Test Positivity Rate*** is the New Case Rate divided by the New Test Rate
  - It is the fraction of tests that come back positive, expressed as a percent.

And some basic nomenclature on vaccination rates

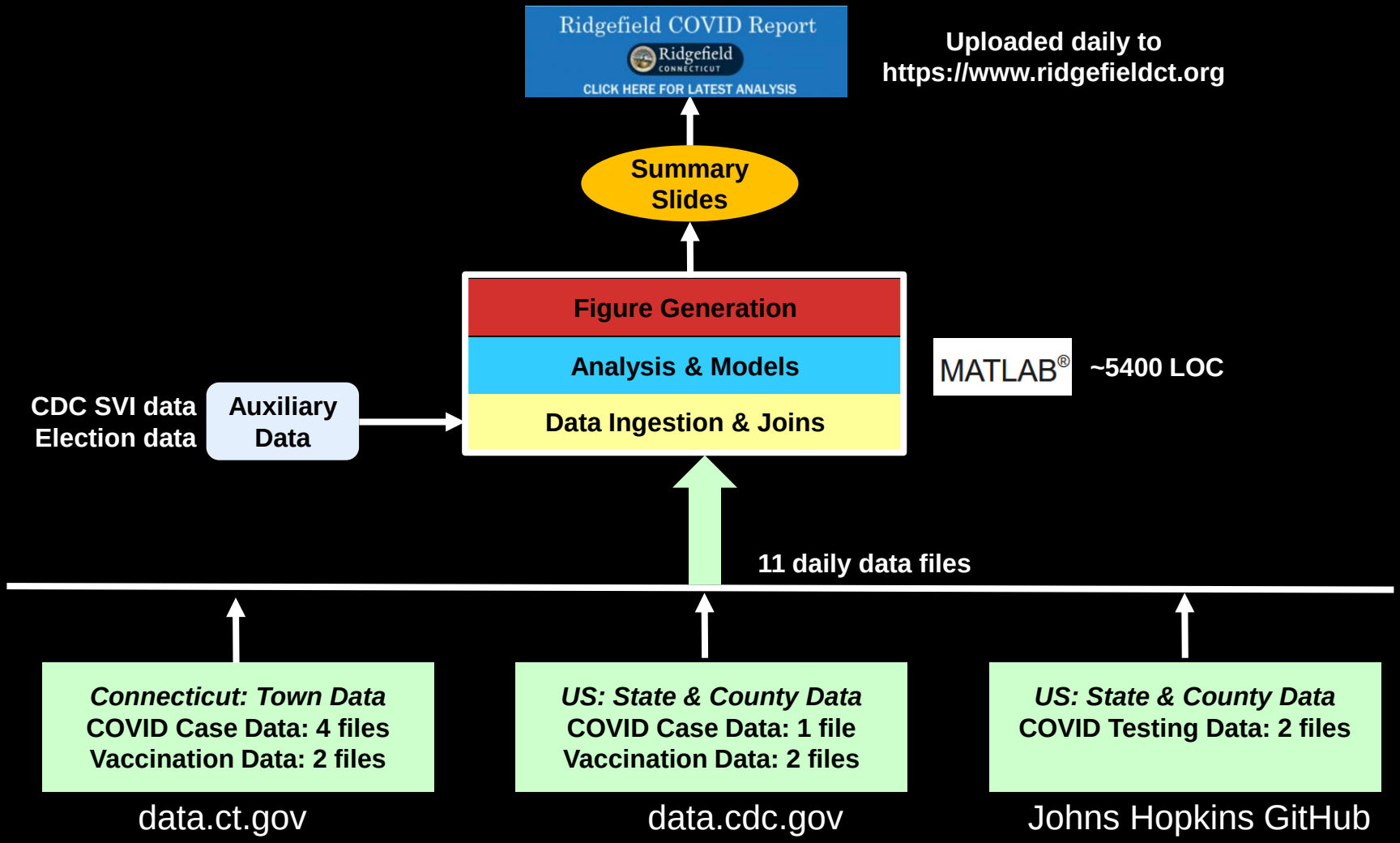
- ***Initiated Vaccination:*** people who have received at least one dose of any vaccine
  - This is sometimes referred to as ***Partially Vaccinated***.
  - It is typically reported as a percent of a specified population group (e.g., Age 12+)

- ***Completed Vaccination:*** people who have competed the recommended number of doses (two for Pfizer or Moderna, one for Johnson & Johnson)
  - This is sometimes referred to as ***Fully Vaccinated***.
  - It is typically reported as a percent of a specified population group (e.g., Age 12+)

- ***Breakthrough Case:*** a person who tests positive more than two weeks after completing vaccination

We will occasionally use the term *Vaccination Rate* to refer to Vaccination Percent (even though, strictly speaking, vaccination percent is not a 'rate' ☺)

We have developed a conventional Data Science framework



COVID-19: Perspectives from a Data Scientist

I. COVID-19 Data Science: a personal journey

II. Introduction to COVID-19 data reporting & data science framework

III. We address the following questions from a pure data-science perspective:

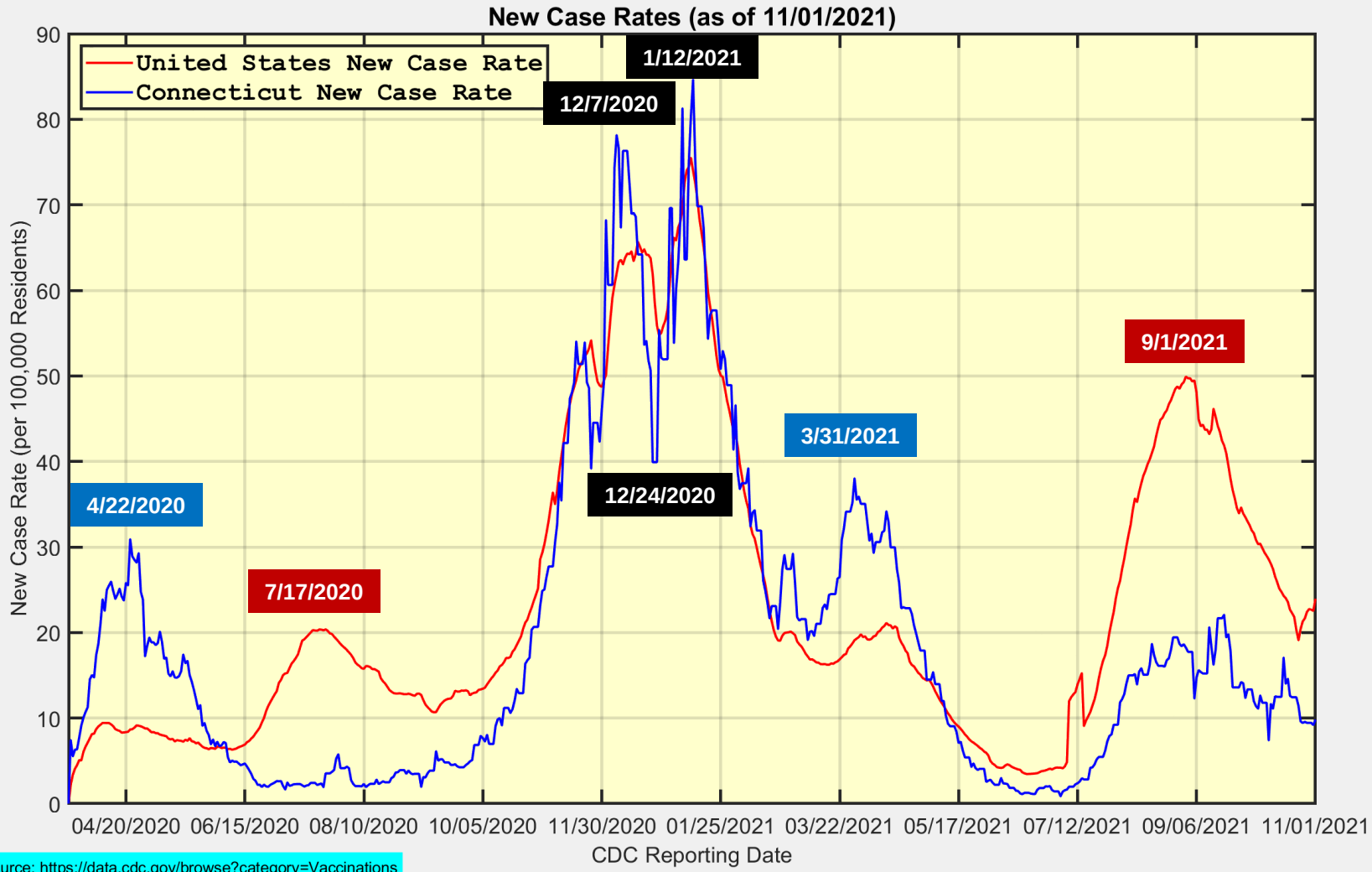
1. What is the current state of the pandemic in the US and Connecticut?
2. Based on the latest vaccination data, will the US reach herd immunity?
3. Do masks reduce COVID infection rates?
4. What is the relationship between vaccination rates, natural immunity, and new case rates?

IV. Personal Perspectives: On the convergence of COVID Data Science and Political Science

Question 1:

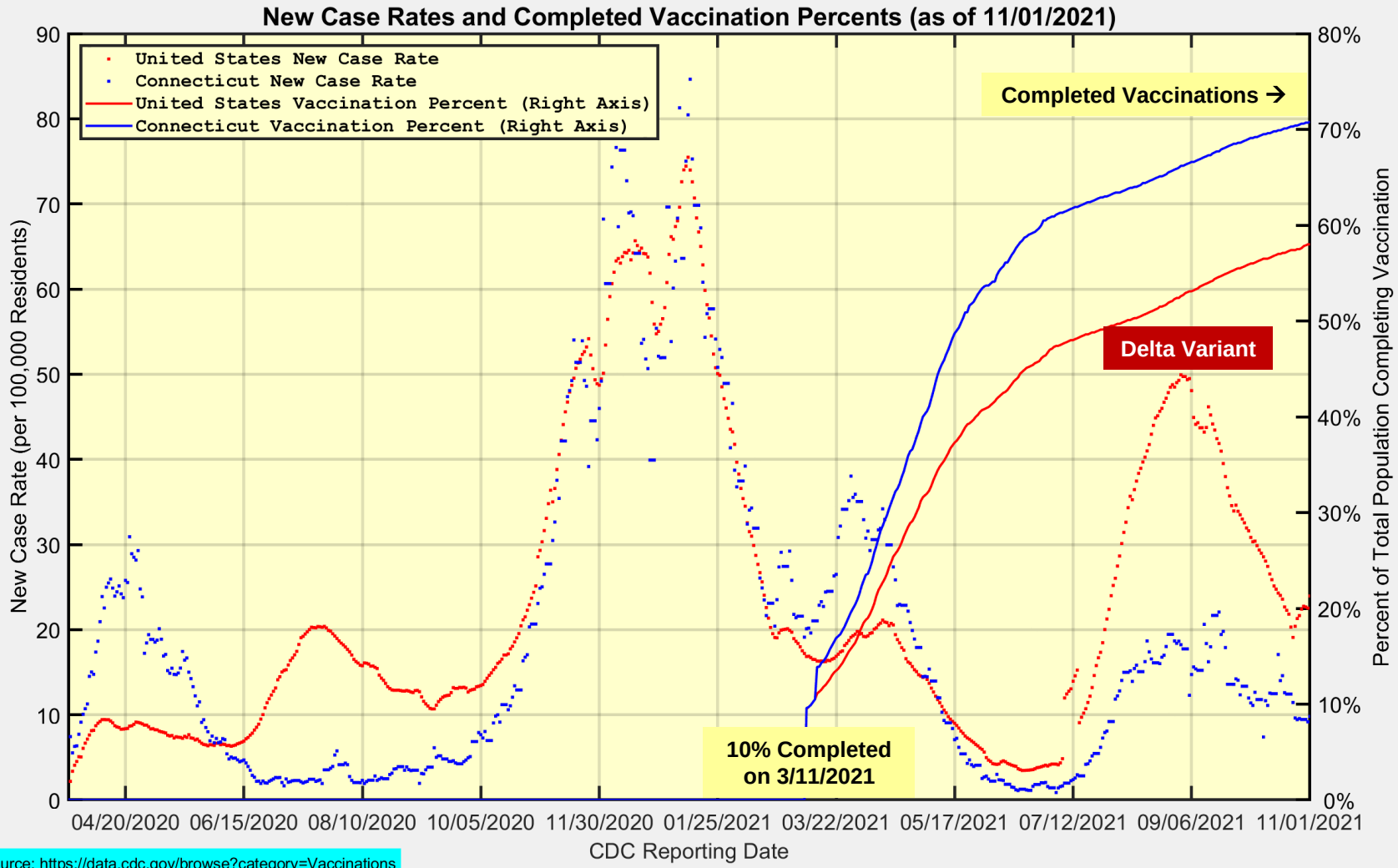
What is the current state of the pandemic in the US and Connecticut?

Progression of the Pandemic: New Case Rates since March 23, 2020



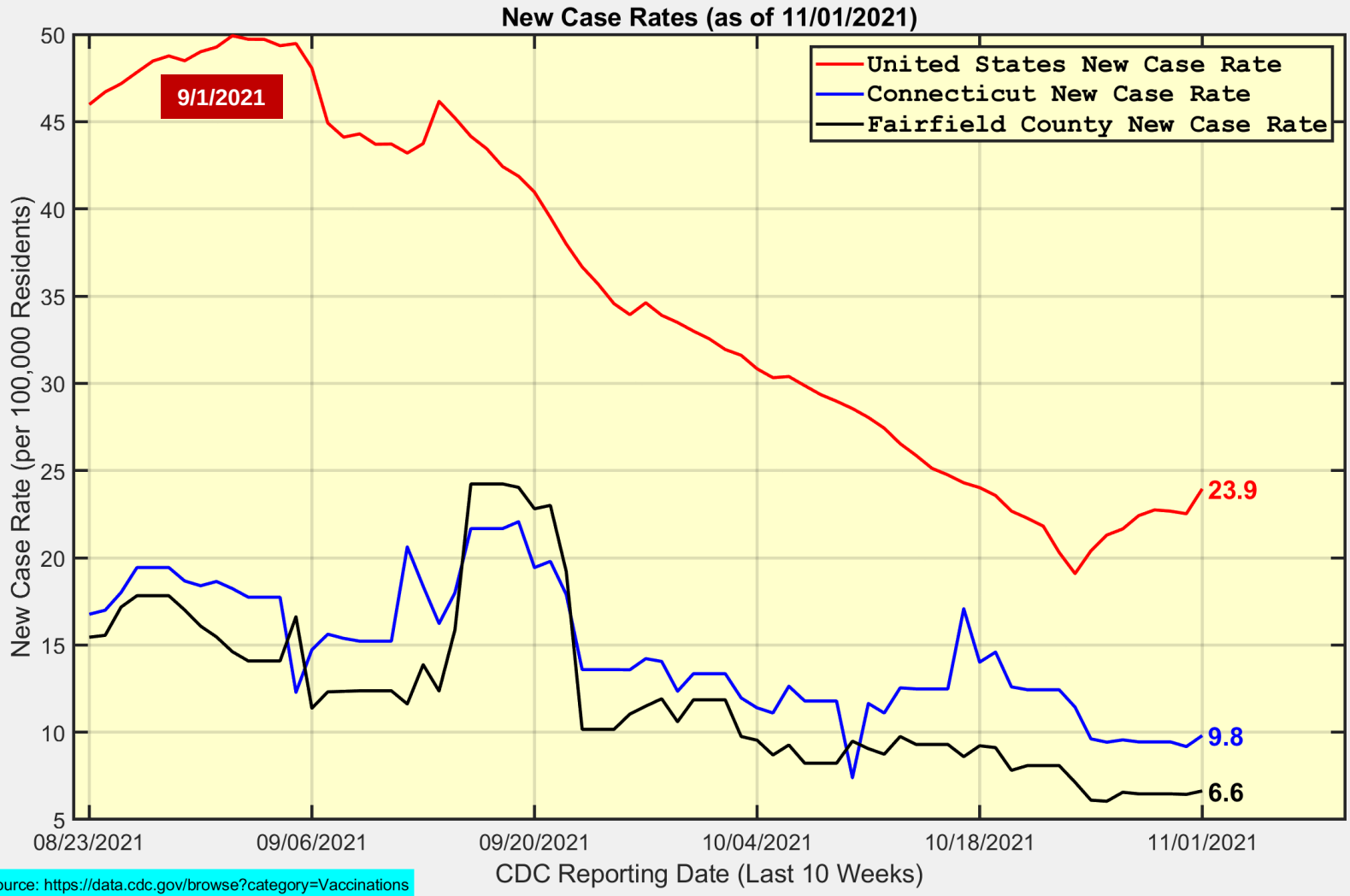
Data Source: https://data.cdc.gov/browse?category=Vaccinations
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

We add Completed Vaccination Percents on the right axis



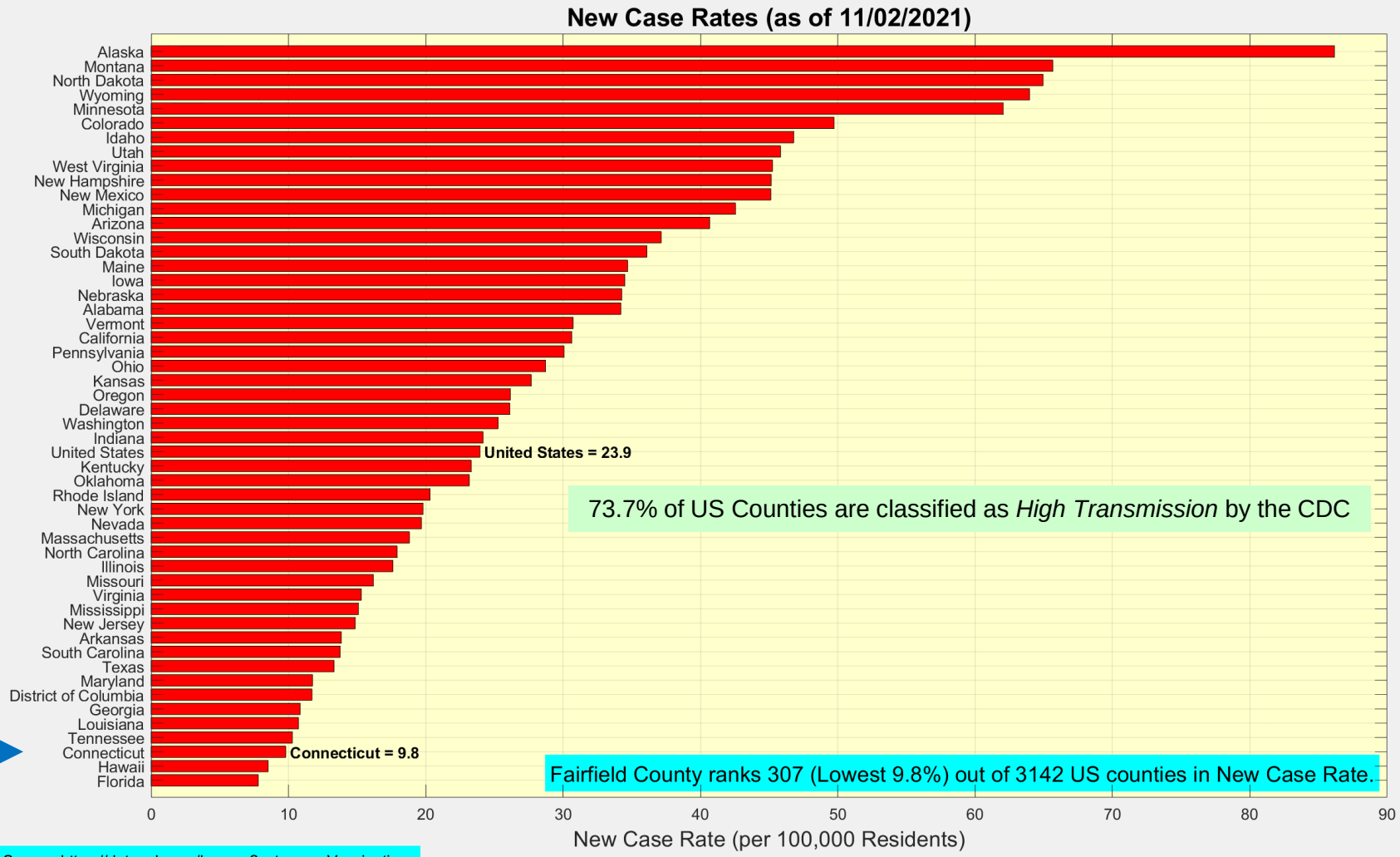
Data Source: https://data.cdc.gov/browse?category=Vaccinations
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

New Case Rates have decreased since September 1



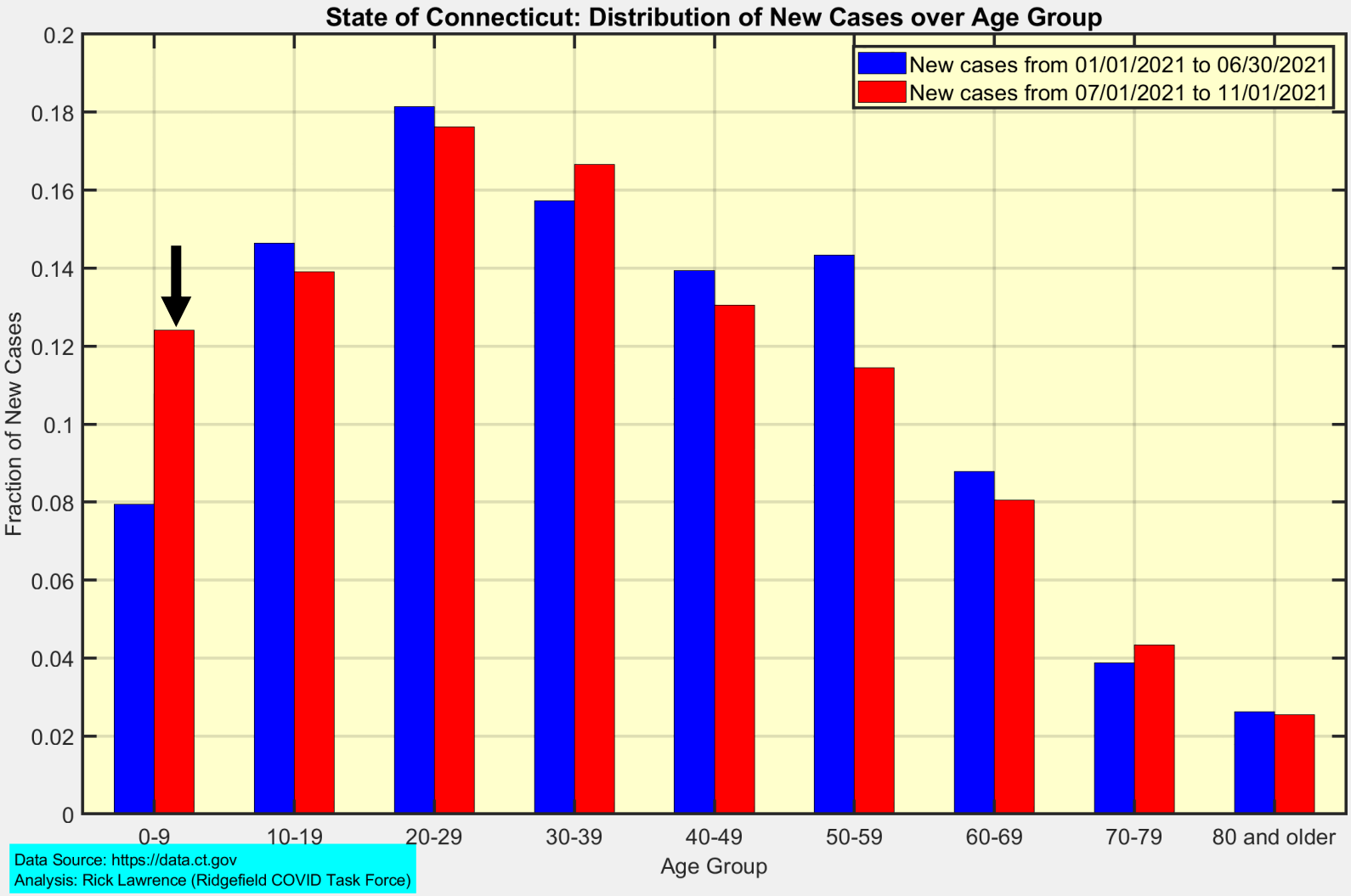
Data Source: https://data.cdc.gov/browse?category=Vaccinations
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

Connecticut has one of the lowest New Case Rates in the US

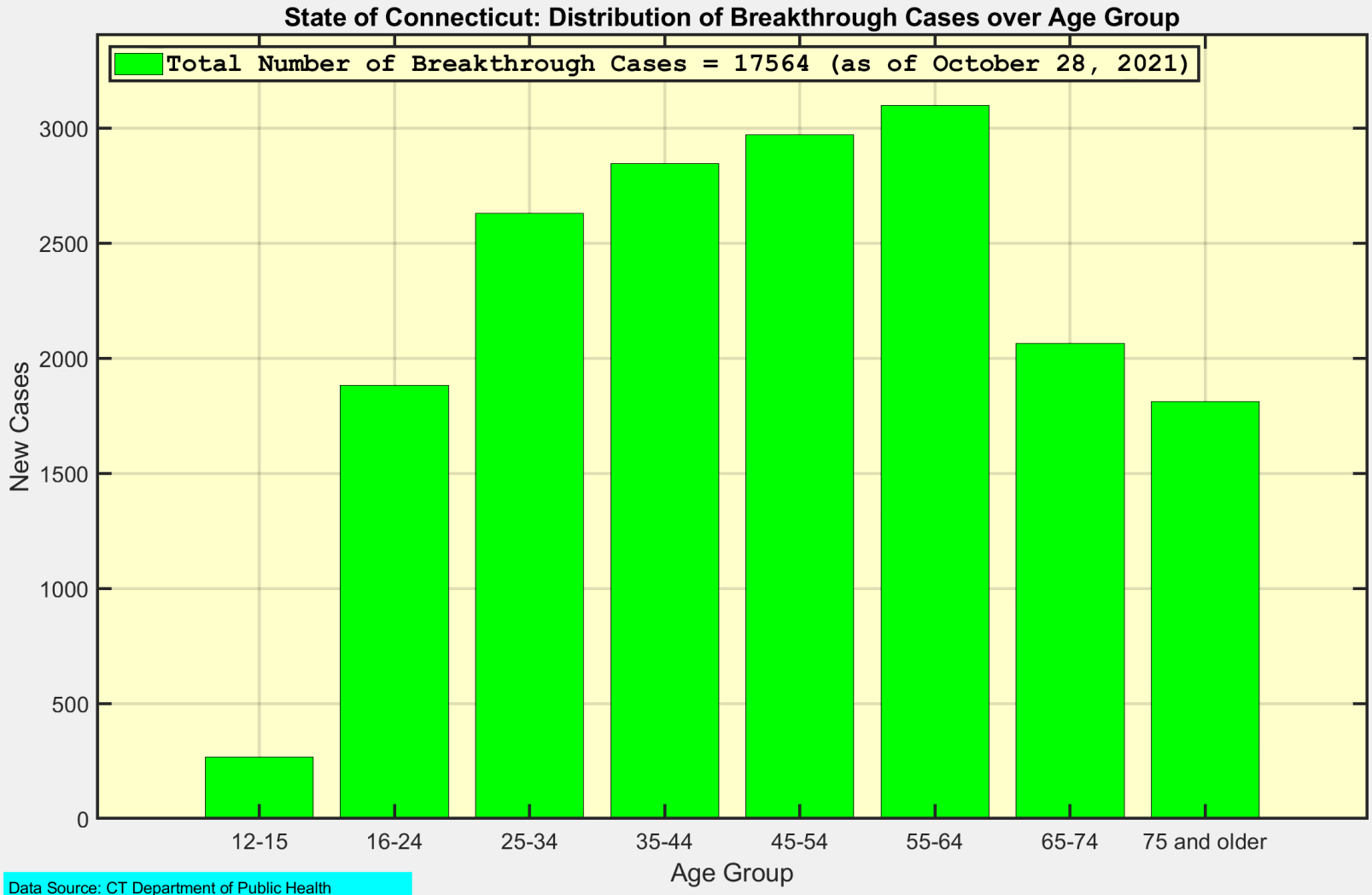


Data Source: https://data.cdc.gov/browse?category=Vaccinations
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

New Cases are occurring more in a younger population



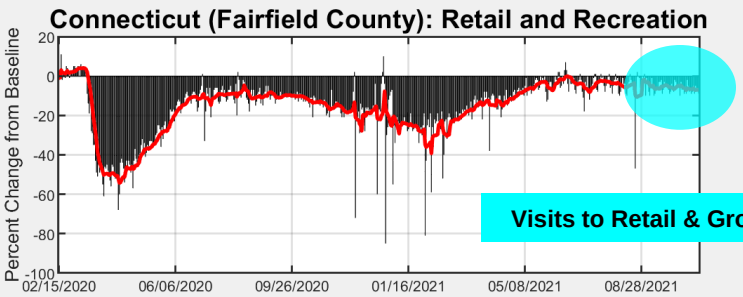
Breakthrough cases occur in higher age groups than all COVID cases

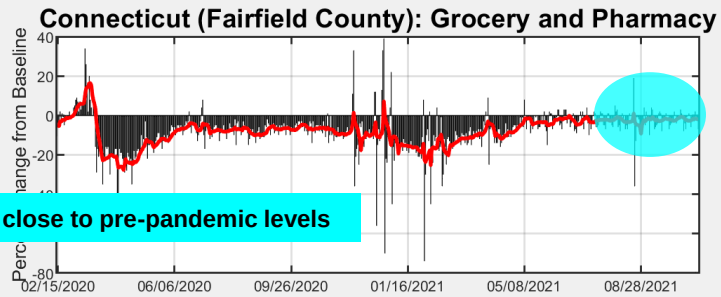


Data Source: CT Department of Public Health
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

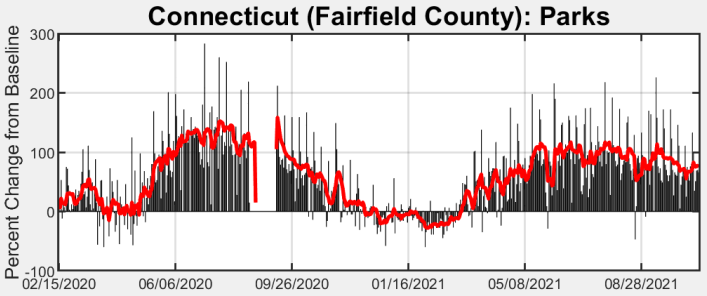
Google Mobility data provides interesting insights into post-pandemic individual behavior

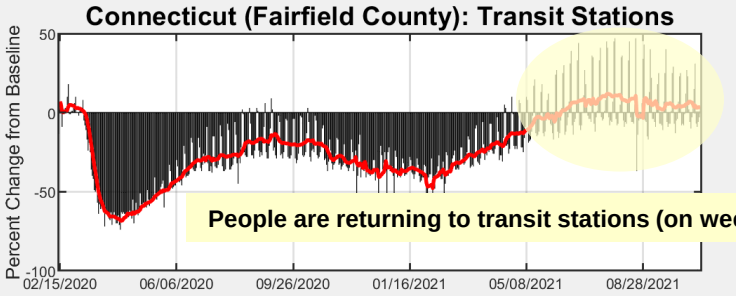
These are percent changes relative to the pre-pandemic baseline (*for Fairfield County*)



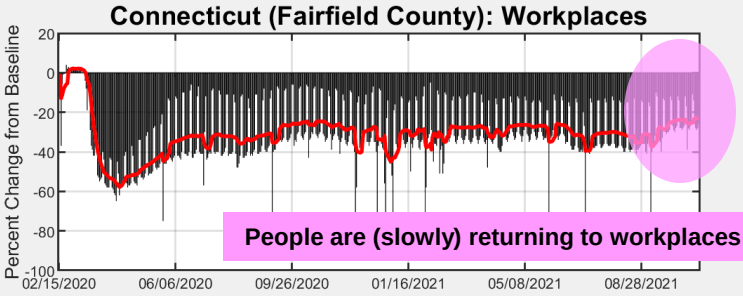


Visits to Retail & Grocery are close to pre-pandemic levels

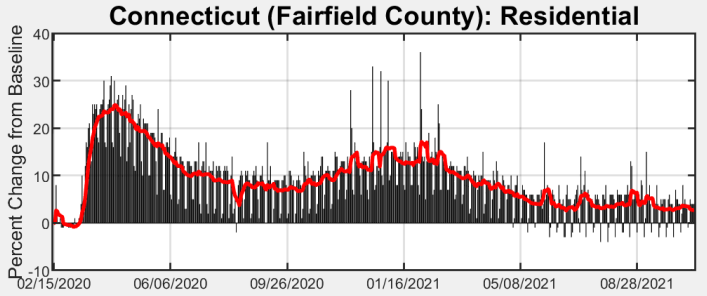




People are returning to transit stations (on weekends)



People are (slowly) returning to workplaces



Data Source: https://www.google.com/covid19/mobility/
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

Question 2:

Based on the latest vaccination data, will the US reach herd immunity?

What is Herd Immunity?

- The *Basic Reproduction Number* (R_0) s the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection

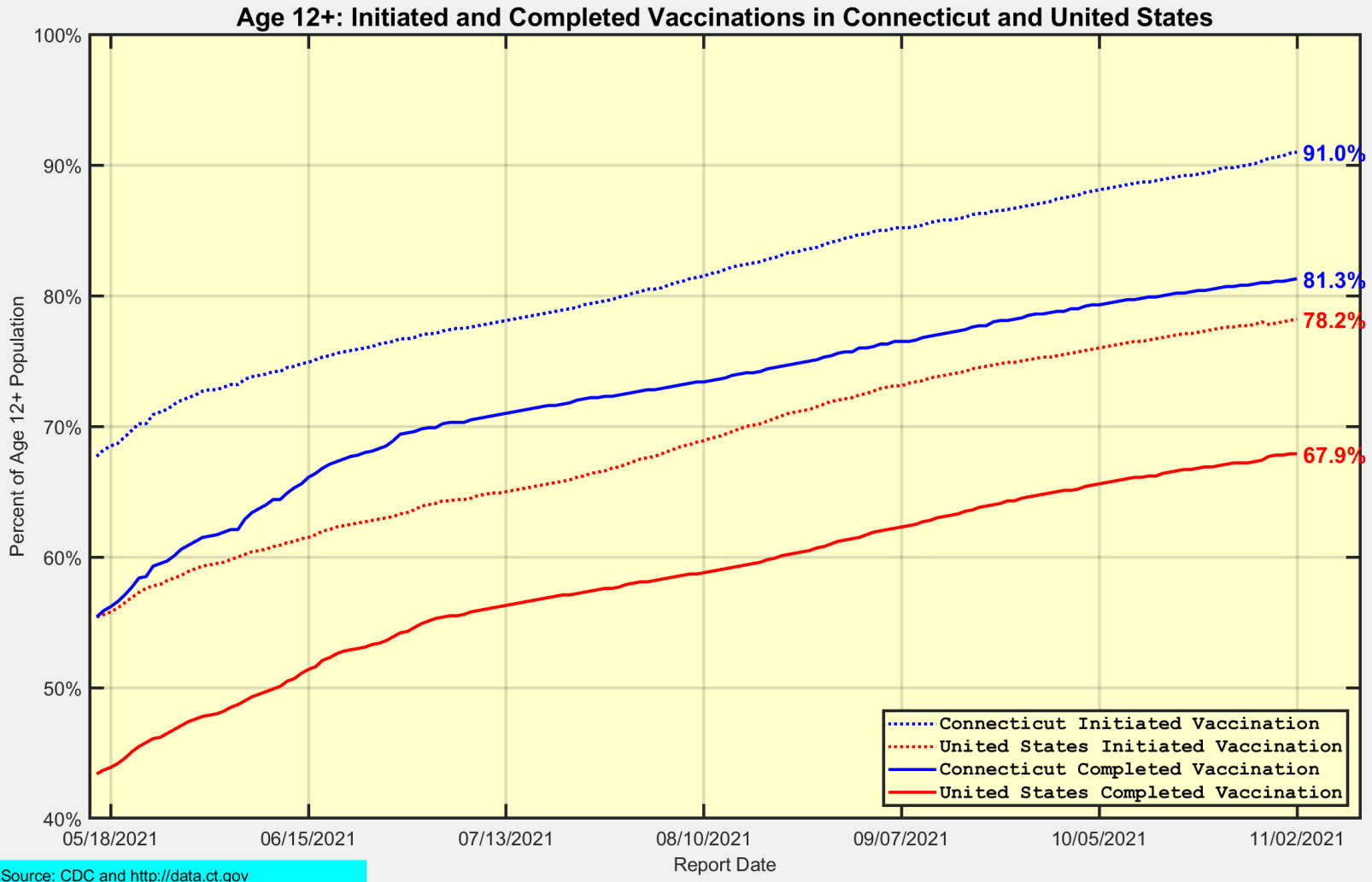
- The *Effective Reproduction Number* (R_t) is the expected number of cases directly generated by one case at time *t* in a partially susceptible population
  - $R_t = R_0 * S$, where S is the fraction of the population that is susceptible to infection

- *Herd immunity* is achieved when a significant fraction of a population acquires immunity, either through vaccination or previous infection (natural immunity)
  - Herd Immunity is achieved when R_t falls below 1.0

- The *Herd Immunity Threshold (HIT)* is the fraction of the population with immunity needed to achieve herd immunity
  - $HIT = 1 – 1/R_0$

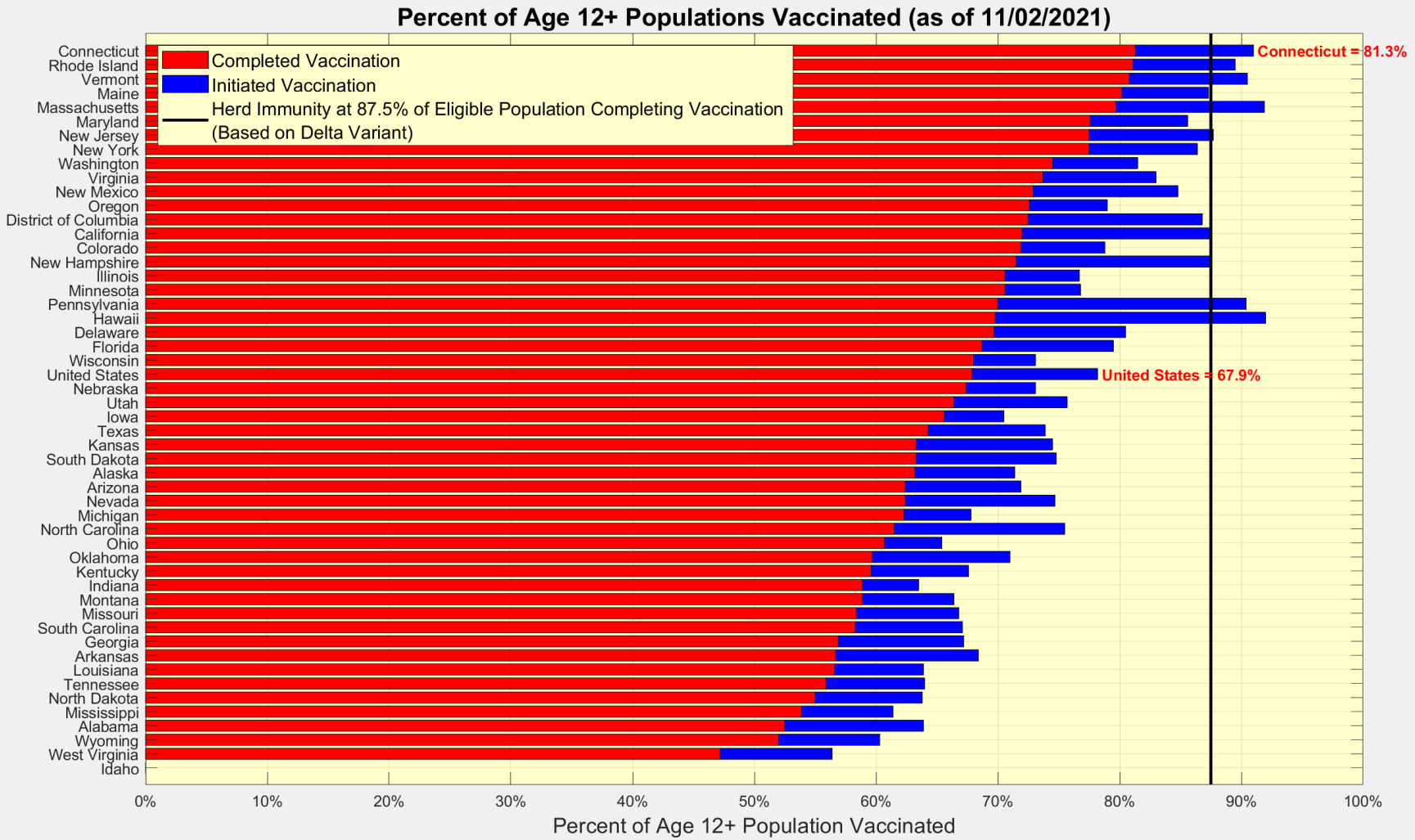
- The Delta variant has an estimated $R_0 = 8$
  - Therefore***, the herd immunity threshold for Delta is 87.5% of the population***

Here are the latest vaccination percents for the eligible (Age 12+) population



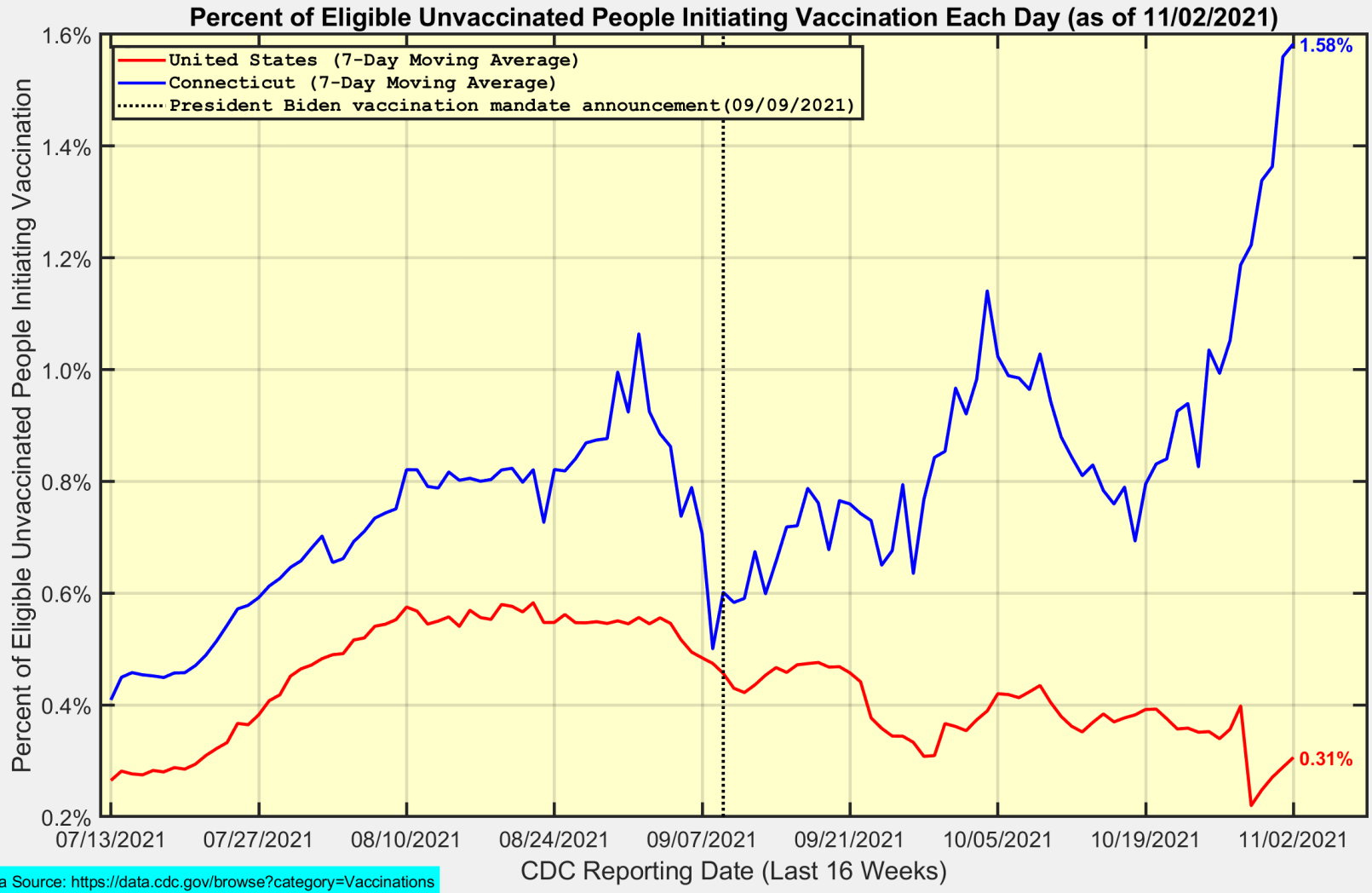
Data Source: CDC and http://data.ct.gov
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

Vaccination percents vary widely across states



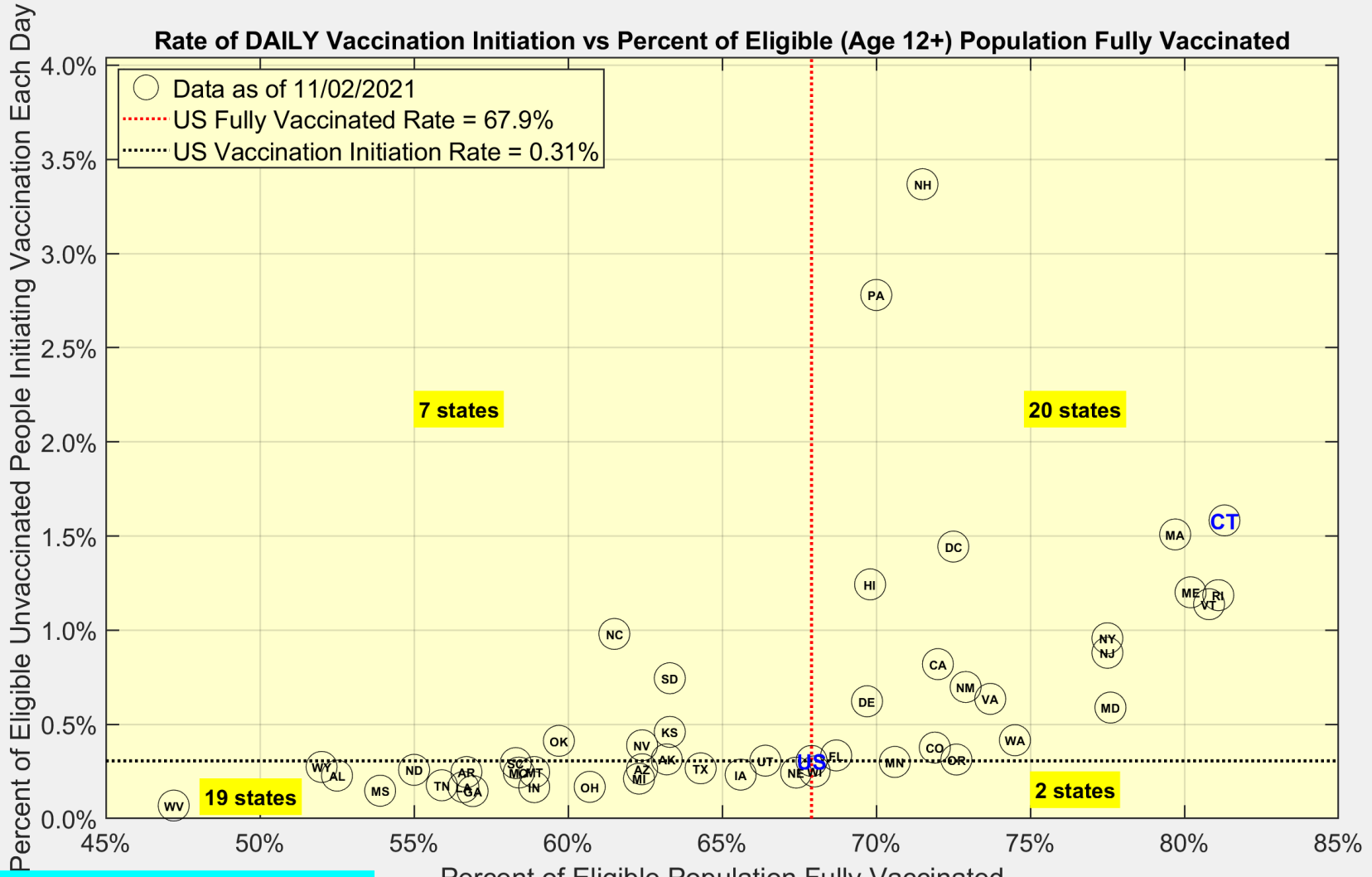
Data Source: https://data.cdc.gov/browse?category=Vaccinations

Important Metric: Percent of Eligible Unvaccinated People Getting First Shot on each day



Data Source: https://data.cdc.gov/browse?category=Vaccinations
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

Major Concern: The states with the lowest vaccination percents are slowing down quicker



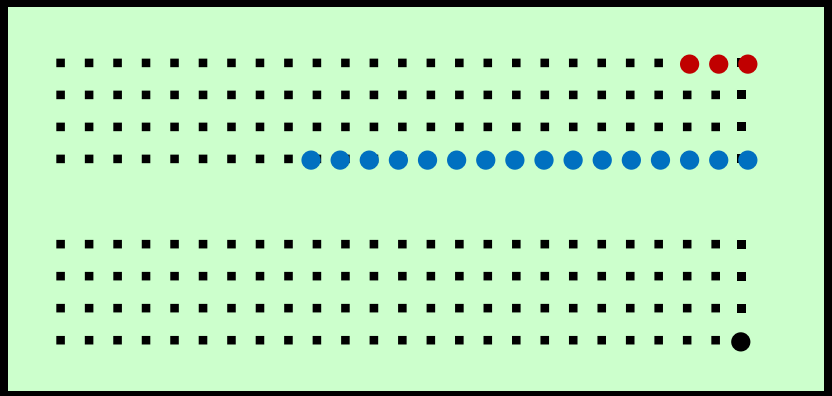
Data Source: https://data.cdc.gov/browse?category=Vaccinations
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

Here is a simpler way to think about this …

- Imagine filling an auditorium each day with a 1000 randomly-selected unvaccinated people who are eligible to get vaccinated

- We then ask the 1000 people if they are willing to initiate vaccination on a given day
  - In the US, approximately **3 people** will raise their hand (on an average day)
  - In Connecticut, approximately **16 people** will raise their hand (on an average day)
  - In West Virginia, only **1 person** will raise their hand

1000 Eligible Unvaccinated People



The United States is unlikely to reach Delta herd immunity based on current vaccination trends … it will require natural immunity (More on that in Question 4 …)

Question 3:

Do masks reduce COVID infection rates?

It is NOT our goal here to review the extensive literature on mask effectiveness …

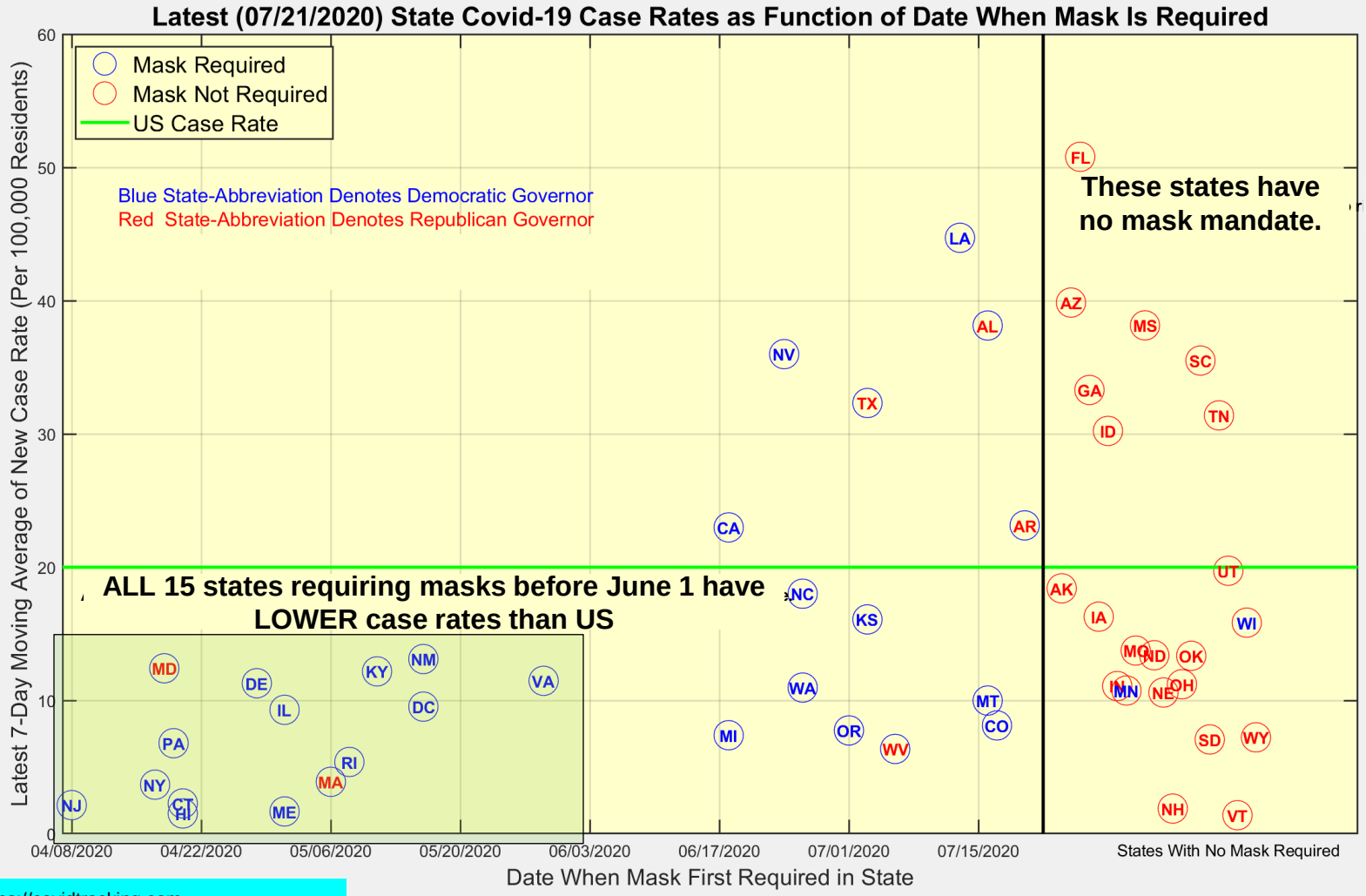
- There have been many studies on the effectiveness of masks against COVID-19
  - These studies range from quasi-clinical 'trials' to observational studies to aerosol dynamics
  - These studies, in whole, confirm that wearing masks significantly reduces COVID-19 transmission

- A review of this extensive literature is beyond the scope of this talk
  - Instead, we report our own analysis of publicly available data from last year

- We quickly note two quasi-clinical 'natural experiments' analyzed by CDC scientists
  - 81 (out of 105) Kansas counties were allowed to opt out of mask mandates in July 2020
  - School districts in Arizona implemented variable mask policies in July & August 2021
  - *CDC analysis found that mask wearing in both settings reduced new cases*

- **Aside:** The 2021 Nobel Prize for Economics was awarded (in part) for methodology in analyzing '*natural experiments*' … there are applications of this work in advanced COVID analysis

This is a slide from *July 21, 2020* when only a subset of states had mask mandates



Data Source: https://covidtracking.com
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

Question 4:

What is the relationship between vaccination rates, natural immunity, and new case rates?

Connecticut breakthrough data shows the dramatic impact of vaccination

1. As of October 28, 2021:
   - There have been 17,564 COVID-19 cases among 2,380,490 fully vaccinated Connecticut residents
   - ***Only 0.74% of vaccinated residents (1 in 135) have contracted the virus***

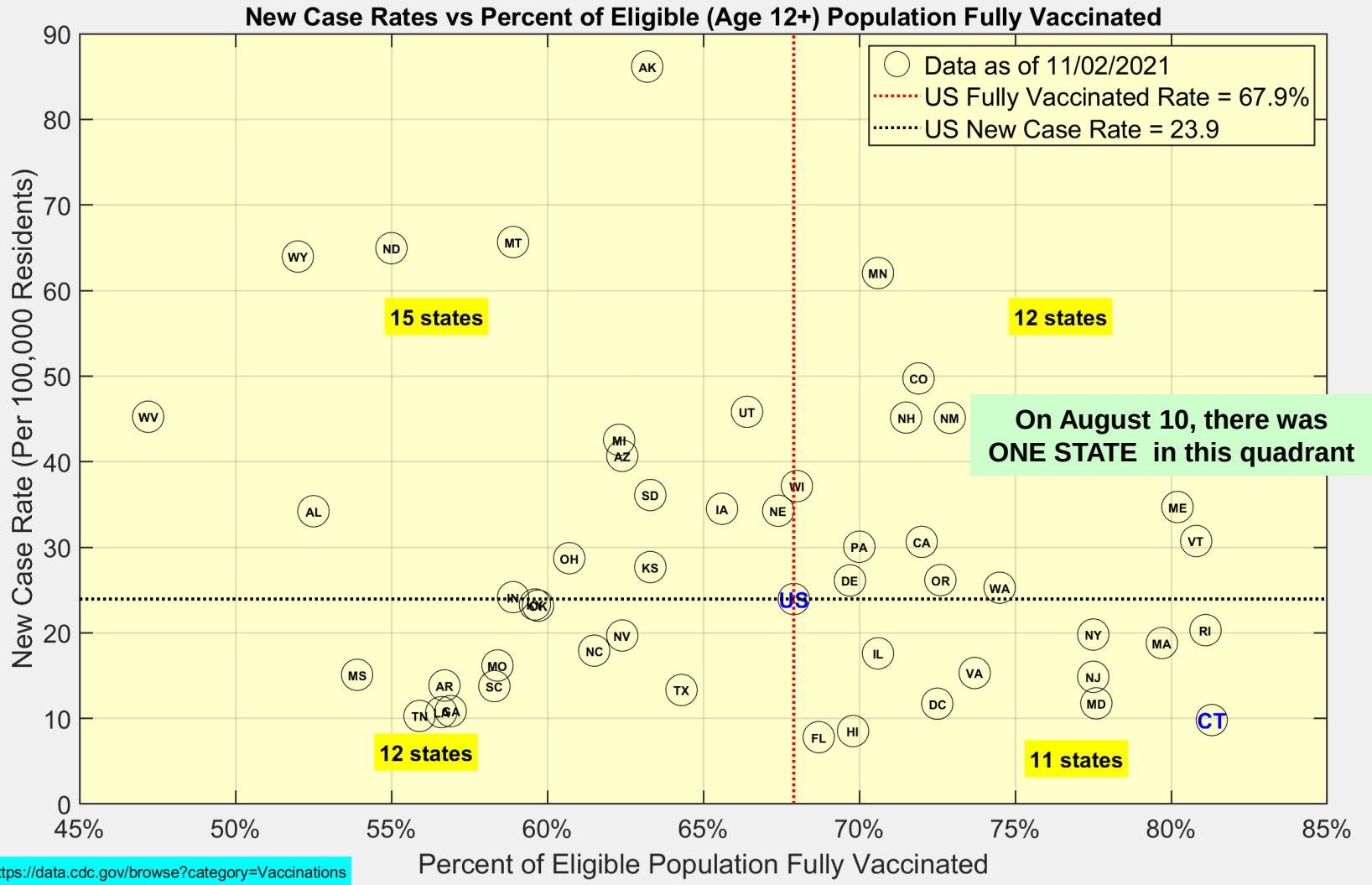
2. Weekly rates as of October 17, 2021:
   - There were 170.1 cases per 100,000 unvaccinated residents
   - There were 34.6 cases per 100,000 fully vaccinated residents
   - ***Unvaccinated residents are 4.9 times more likely to get COVID-19 than vaccinated residents***

   - There were 47.7 hospitalizations per 100,000 unvaccinated residents
   - There were 2.6 hospitalizations per 100,000 fully vaccinated residents
   - ***Unvaccinated residents are 18.4 times more likely to require hospitalization than vaccinated residents***

   - There were 4.16 deaths per 100,000 unvaccinated residents
   - There were 0.23 deaths per 100,000 fully vaccinated residents
   - ***Unvaccinated residents are 18.1 times more likely to die from COVID-19 than vaccinated residents***

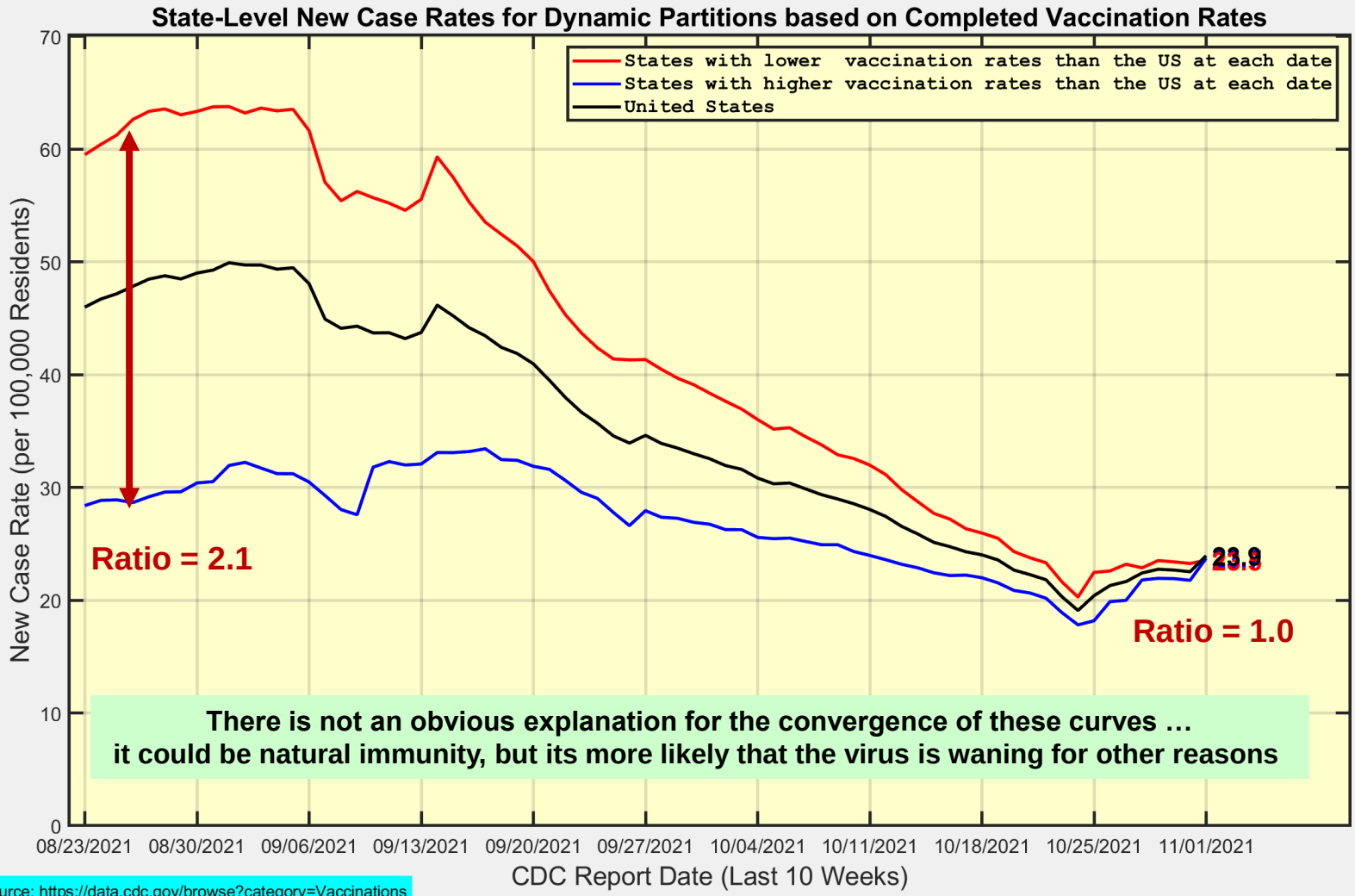
Data Source: Connecticut Department of Public Health

States with lower vaccination fractions have higher New Case rates



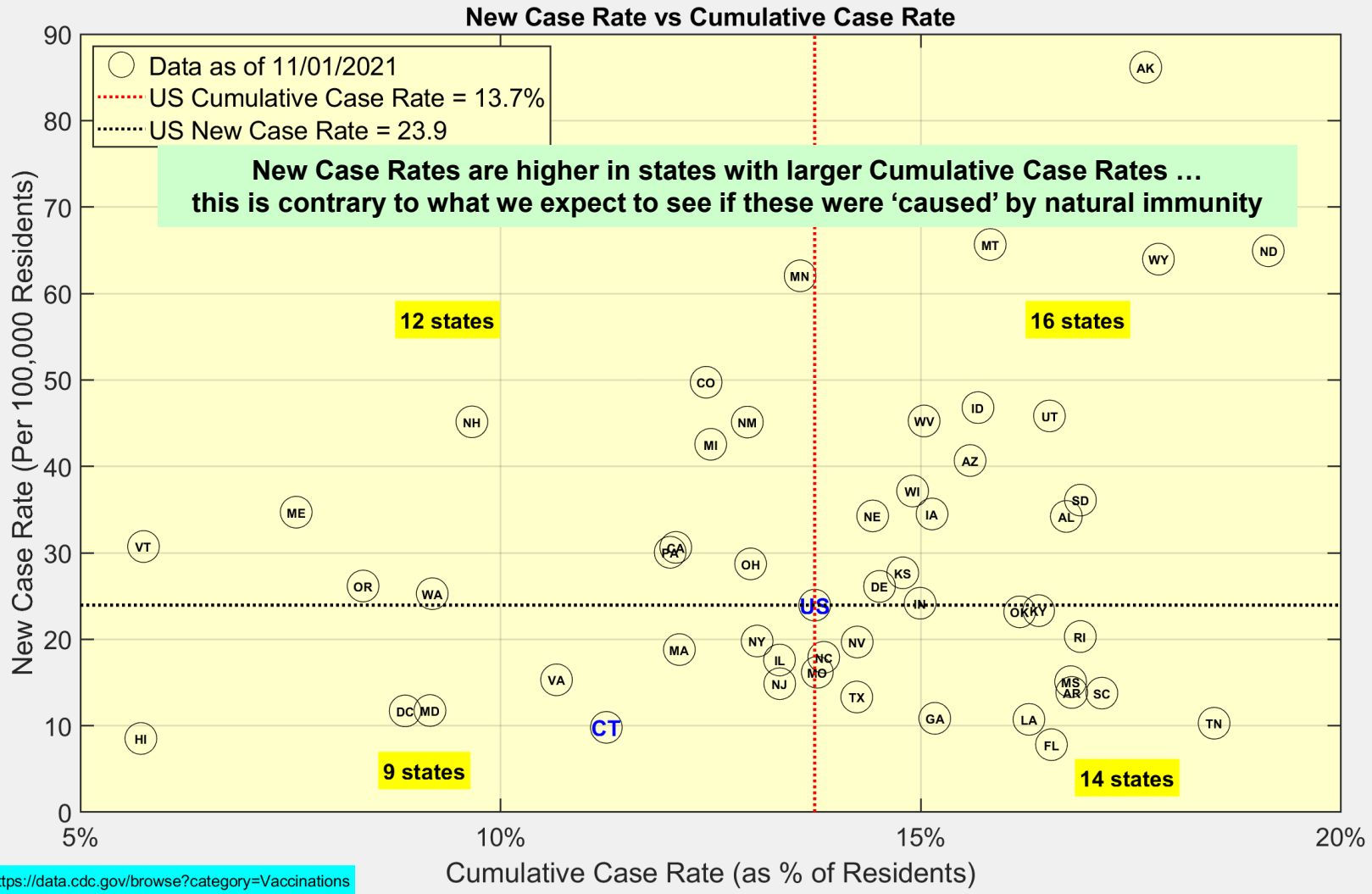
Data Source: https://data.cdc.gov/browse?category=Vaccinations
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

Higher vaccinated states have had lower New Case Rates than lesser vaccinated states



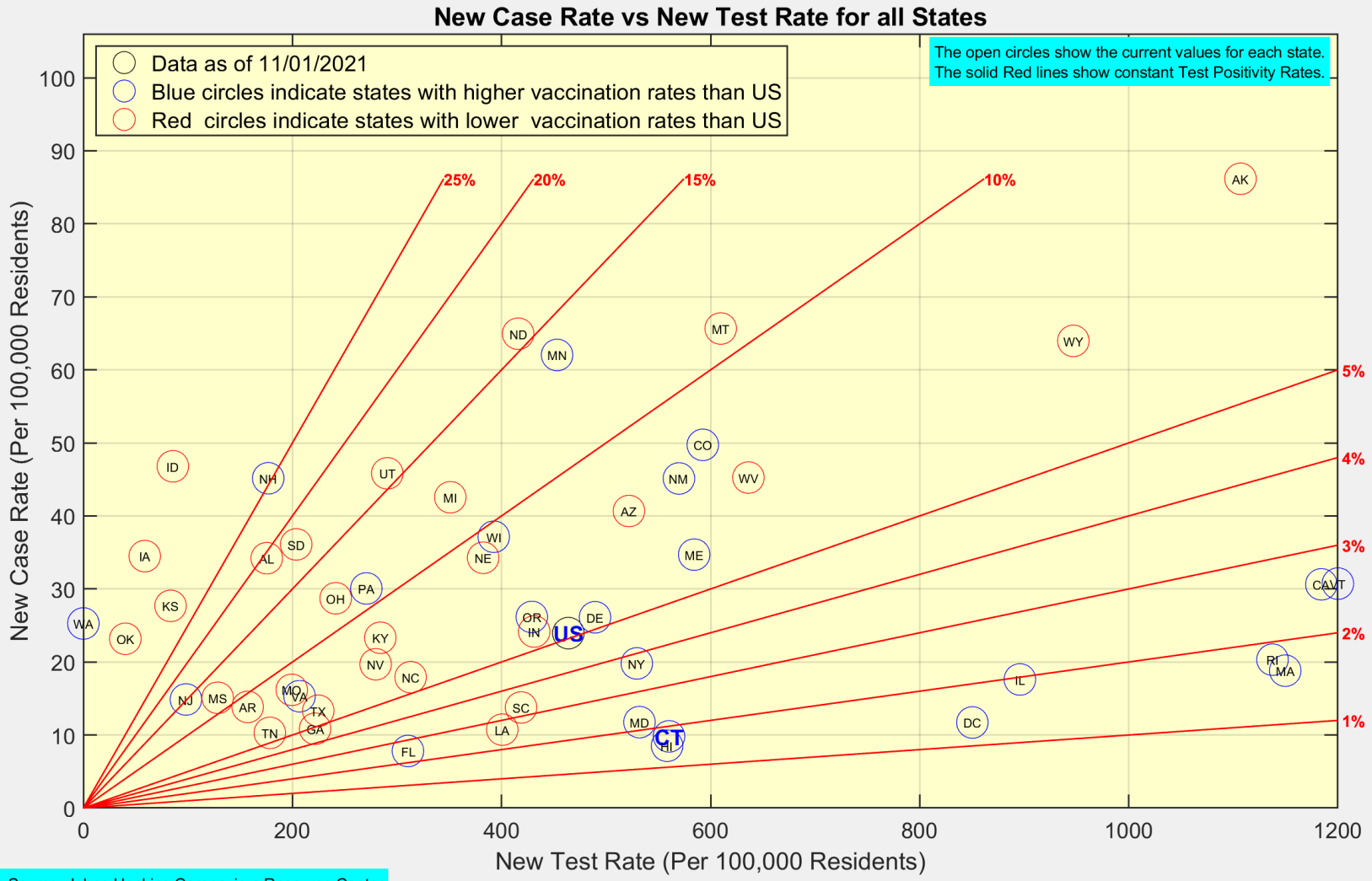
Data Source: https://data.cdc.gov/browse?category=Vaccinations
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

1. Natural Immunity (from previous infection) does NOT appear to lead to reduced new cases



Data Source: https://data.cdc.gov/browse?category=Vaccinations
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

2. Less vaccinated states are doing less testing than higher vaccinated states



Data Source: Johns Hopkins Coronavirus Resource Center
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

We use Lasso Regression Models to help explain the state-level data

1. Lasso is a class of statistical techniques that mathematically constrains the model to require fewer parameters and hence yield simpler models with greater explanatory predictive power
   - Feature selection is done automatically by the algorithm such that overfitting is minimized

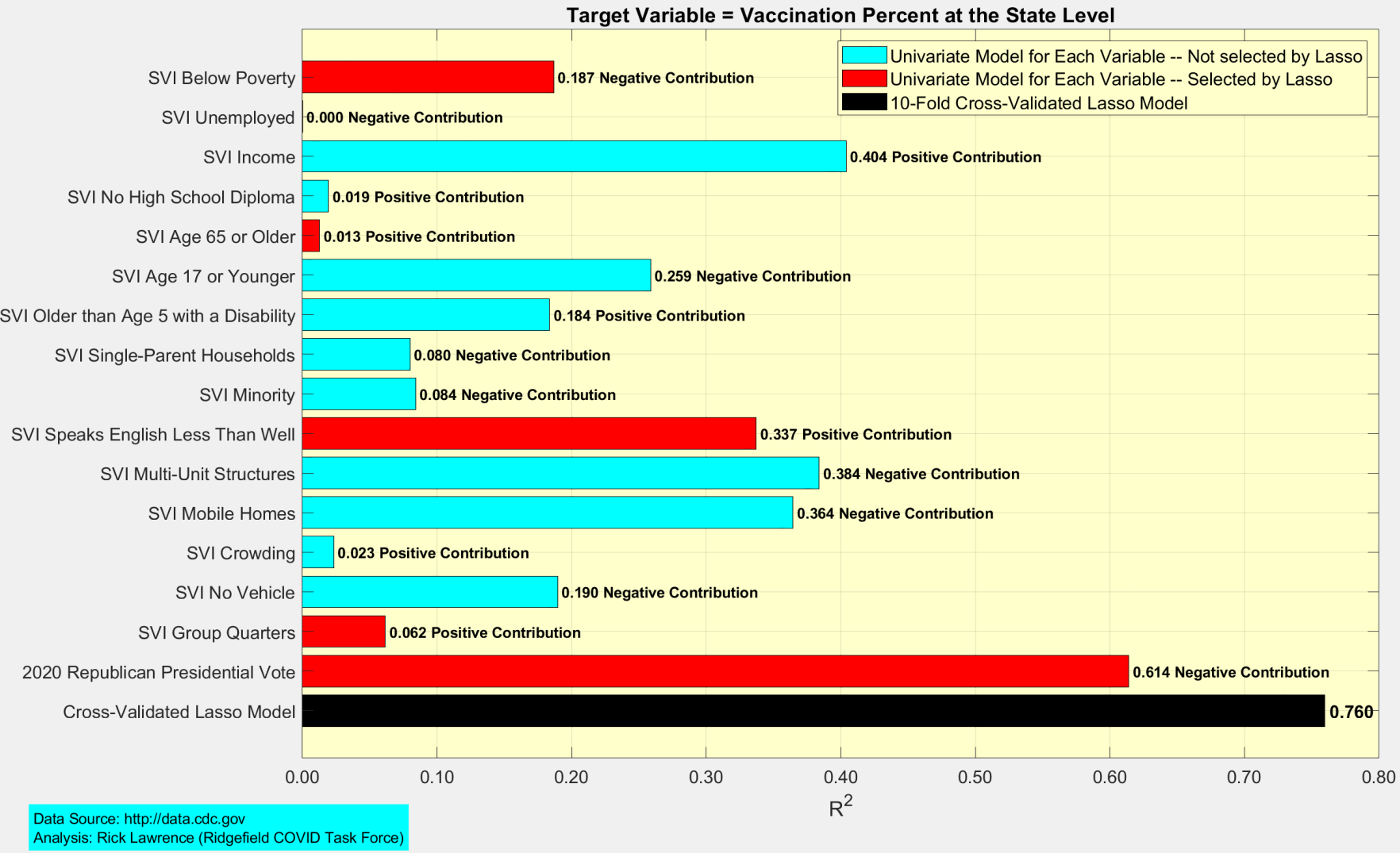
2. We use Lasso with two different target (prediction) variables
   1. Vaccination rates at the state level
   2. New Case Rates (over the past 60 days) at the state level

3. We use CDC ***Social Vulnerability Index (SVI)*** metrics as explanatory (input) variables
   - The CDC has published 15 SVI metrics that help officials identify communities requiring assistance before, during, and after a natural disaster
   - These metrics measure *socio-economic status, household composition, minority status, and housing status*

4. We add additional metrics as explanatory variables
   - 2020 Republican Presidential Vote
   - Cumulative number of cases (to predict New Case Rates) as a proxy for Natural Immunity
   - Vaccination rates (to predict New Case Rates) as measure of Vaccine-based Immunity

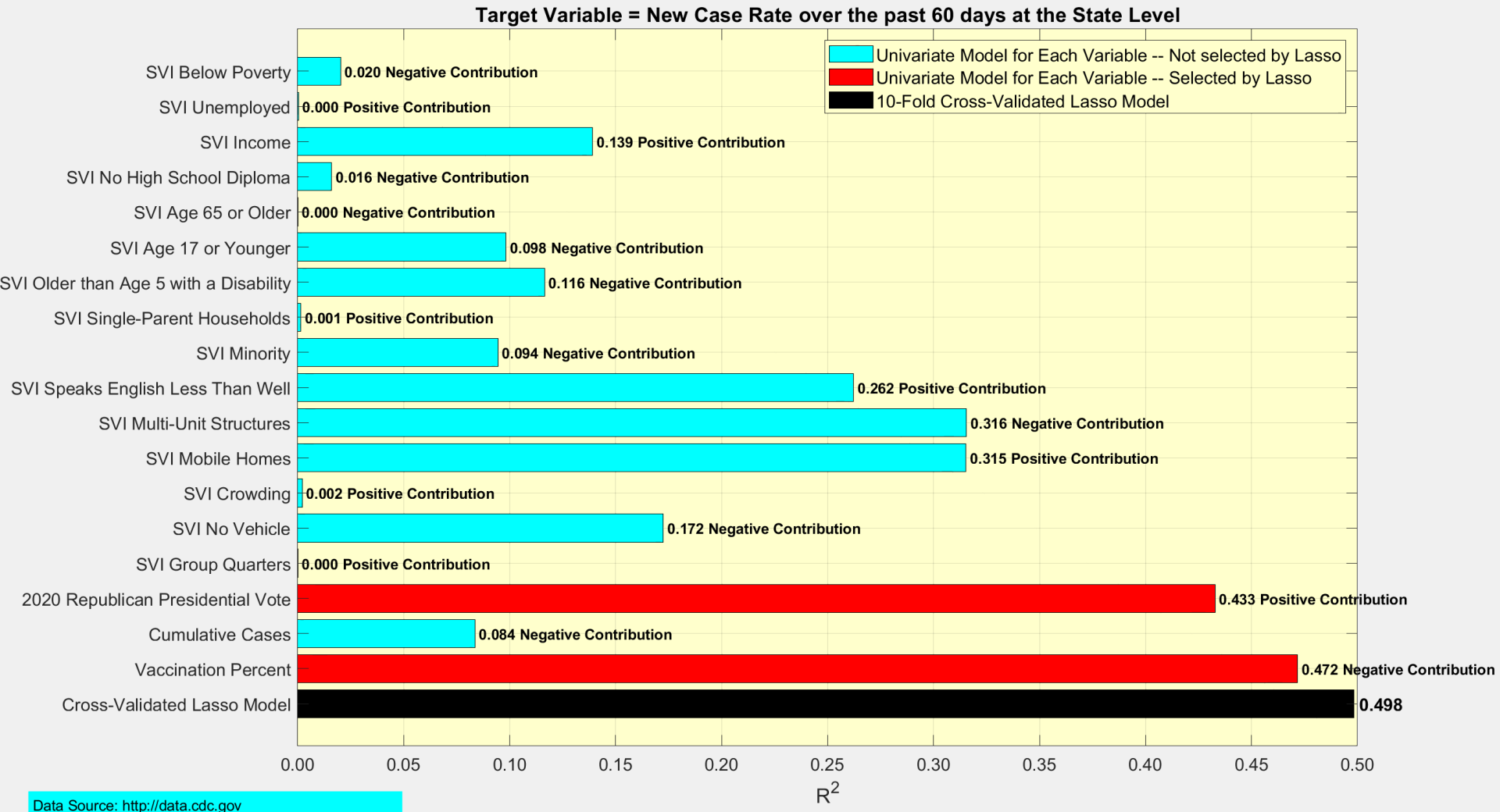
5. Our objective is to *explain* (quantitatively) Vaccination Rates and New Case Rate in terms of this set of explanatory variables

Lasso Model: Target Variable = Vaccination Percent at the State level



Data Source: http://data.cdc.gov
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

Lasso Model: Target Variable = New Case Rate at the State level



Data Source: http://data.cdc.gov
Analysis: Rick Lawrence (Ridgefield COVID Task Force)

Summary of Lasso Models

Model 1: Vaccination Percent at the State Level

Target Variable = Vaccination Percent at the State Level			
Explanatory Variable	**Contribution**	**R2 Value**	**Observation**
Full Lasso Model		**0.760**	76% of the variance in Vaccinate Rate is captured by the model
2020 Republican Presidential Vote	Negative	**0.614**	**Vaccination Rates are lower in 2020 Republican voting states**
SVI Speaks English Less Than Well	Positive	0.337	*Vaccination Rates are higher where less English is spoken*
SVI Below Poverty	Negative	0.187	Vaccination Rates are lower where there is higher poverty
SVI Group Quarters	Positive	0.062	Vaccination Rates are higher where people live in group living
SVI Age 65 or Older	Positive	0.013	Vaccination Rates are higher where there are more 65+ residents

Model 2: New Case Rate at the State Level

Target Variable = New Case Rate at the State Level			
Explanatory Variable	**Contribution**	**R2 Value**	**Observation**
Full Lasso Model		**0.498**	50% of the variance in New Case Rate is captured by the model
Vaccination Percent	Negative	**0.472**	**New Case Rates are lower in states with higher vaccination rates**
2020 Republican Presidential Vote	Positive	0.433	New Case Rates are higher in 2020 Republican voting states

Other Model-based findings:

- The number of previous cases (i.e., Natural Immunity) is not a significant predictor of New Case Rates

COVID-19: Perspectives from a Data Scientist

I. COVID-19 Data Science: a personal journey

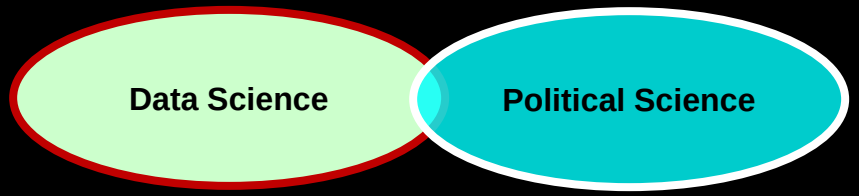
II. Introduction to COVID-19 data reporting & data science framework

III. We address the following questions from a pure data-science perspective:
1. What is the current state of the pandemic in the US and Connecticut?
2. Based on the latest vaccination data, will the US reach herd immunity?
3. Do masks reduce COVID infection rates?
4. What is the relationship between vaccination rates, natural immunity, and new case rates?

IV. Personal Perspectives: On the convergence of COVID Data Science and Political Science

From COVID Data Science to Political Science: A Personal Perspective

So far, we have strived to stay within the boundaries of Data Science.
But we now examine the convergence with Political Science.
By design, this is a personal perspective.



- A simple 'proof':
  1. Data science confirms that masks and vaccinations significantly reduce the risk to individuals of contracting COVID-19. *(And 424M US vaccine doses to date confirm minimal risk of serious side effects.)*
  2. Governments have an obligation to ensure the safety of their constituents by minimizing known risks.
  3. Therefore, government-instituted mask and vaccination mandates do not overstep governmental authority.

- And a stronger corollary:
  4. Given the known data science, Governments would fail in its obligations to constituents if they did NOT mandate the use of masks and vaccinations to mitigate societal risk.

But what about personal freedom?

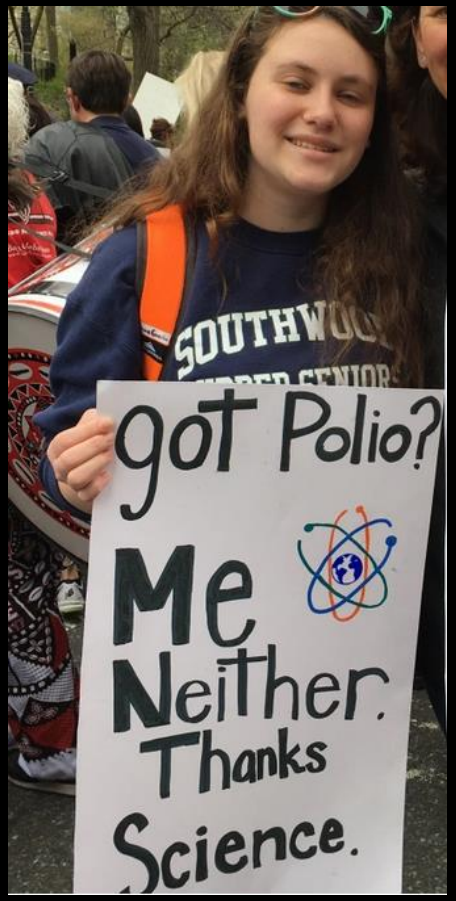
1. Automobile safety is relevant here
   a. We accept that government has a right to mandate the use of seatbelts because extensive data has shown that seatbelts provide protection to people who use them
   b. We accept that government has a right to enforce drunk-driving laws because extensive data has shown that drunk drivers kill innocent people

2. This can be viewed from a risk perspective
   a. If you choose to ignore seatbelt laws, you accept the personal risk of increased injury in an accident
   b. If you choose to drive drunk, you are passing the risk of severe injury (or death) to innocent people

3. Why is COVID-19 any different?
   a. I will (reluctantly) acknowledge that you have the freedom to accept the personal risk associated with not wearing a mask or remaining unvaccinated
   b. But I do not accept that you have the freedom to transfer that risk to me (by increasing my chances of getting infected)
   c. Nor will I accept that you have the freedom to transfer that risk to society (via increased healthcare costs, broader economic impact, and increased suffering)

Science has a long history of improving our lives … why reject it now?

It's useful to remember that vaccines have eradicated Polio (and Smallpox)



Science has given us multiple high-efficacy Covid-19 vaccines.
We have the tools to end this pandemic.

THANK YOU

Rick Lawrence
ricklawrence000@gmail.com
https://www.linkedin.com/in/Rick-Lawrence/