

COVIDMINDER Example Notebook

DAR/HACL Summer 2020

John S. Erickson

07/13/2020

INTRODUCTION

This R Notebook and its related R scripts provide a very basic introduction to the **COVIDMINDER** application, which you can find at:

- Public “production” version: <https://covidminder.idea.rpi.edu/> (<https://covidminder.idea.rpi.edu/>)

The github repository for all the code required for this notebook, including a snapshot of the **COVIDMINDER** application, can be found at:

- <https://github.com/TheRensselaerIDEA/COVID-Notebooks> (<https://github.com/TheRensselaerIDEA/COVID-Notebooks>)

The **COVIDMINDER** github repository can be found at:

- <https://github.com/TheRensselaerIDEA/COVIDMINDER>
(<https://github.com/TheRensselaerIDEA/COVIDMINDER>)

NOTE: To contribute or submit to these repositories you must have a github.com ID and must send it to John Erickson at `erickj4@rpi.edu`. **PLEASE DO THIS IMMEDIATELY BEFORE READING ANY FURTHER!!**

HOW TO USE THIS NOTEBOOK AND REPOSITORY

In this exercise we're asking you to...

- clone the COVIDMINDER Notebook github repository...
- create a personal branch, and...
- make additions to the repository by creating new, customized notebooks.

The instructions which follow explain how to accomplish this.

For DAR/HACL Summer 2020 you *must* be using RStudio Server on the IDEA Cluster. Instructions for accessing “The Cluster” appear at the end of this notebook. Don’t forget to send your github ID to `erickj4@rpi.edu`

The general procedure for cloning and using this repository is as follows:

- Access the Linux shell on the IDEA Cluster by clicking the **Terminal** tab of RStudio Server or by using `ssh` to directly access the Cluster. `cd` into your home (login) directory.
- `git clone https://github.com/TheRensselaerIDEA/COVID-Notebooks.git` in your home directory, resulting in a new directory `COVID-Notebooks`
- In the Linux shell, `cd` to `COVID-Notebooks`
- In the Linux shell, `git checkout -b hacl-yourrcs` where `yourrcs` is your RCS id
 - If your RCS is `erickj4`, your new branch should be `hacl-erickj4`

- It is critical that you include your RCS id in your branch id
- In RStudio Server, navigate to COVID-Notebooks via the **Files** panel
- Set this to be your R working directory (under “More” menu)
- Recommended procedure: Make a **copy** of this notebook (ie `Rmd`) file using an *original, descriptive* filename that **includes your RCS ID!**
- Alternate procedure: Create an entirely new, fresh notebook.
- In either case: use the data we’ve provided under the `data/csv` subdirectory.
- Edit the notebook...`knit` (to HTML)...repeat
- In Linux, `git add` each file you want to add to the repository (e.g. your new `Rmd` file, perhaps the `html` you create when you `knit`)
- When you’re ready, in Linux `git commit -a -m "some comment"` where “some comment” is a useful comment describing your changes
- Finally, `git push origin hacl-yourrcs` (where `hacl-yourrcs` is the branch you’re working on)
- ...and then go to github, select your branch from the drop-down and submit a pull request.

Please also see this handy github “cheatsheet”: [\(https://education.github.com/git-cheat-sheet-education.pdf\)](https://education.github.com/git-cheat-sheet-education.pdf)

UNDERSTANDING THE DATA

- All of the data used for this notebook and the **COVIDMINDER** app is located in the `data/` directory hierarchy, mostly under `data/csv/` and `data/csv/time_series/`
- Feel free to browse the directory tree using RStudio or on the web via github
- Most of the nationwide data is under `data/csv`
- Most of the nationwide time series data and New York per-county data is under `data/csv/time_series/`
 - This layout may change
- We will be expanding our data archive, esp. using social determinants data from **County Health Rankings**
 - See [\(https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/2020-measures\)](https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/2020-measures)

WHAT SHOULD I WORK ON?!?!

- Choose an existing issue: [\(https://github.com/TheRensselaerIDEA/COVID-Notebooks/issues\)](https://github.com/TheRensselaerIDEA/COVID-Notebooks/issues)
 - Ask questions about the issue in the “Comments” section of that issue
 - “Claim” the issue by self-assigning
- OR, create your own issue; click “New Issue”
- OR, ask Prof. Bennett or Dr. Erickson for an issue to work on!

HOW TO SHARE YOUR WORK OR ASK QUESTIONS

- For class-related questions, use the `data-incite-2020` slack channel under The Rensselaer IDEA
- For COVIDMINDER questions use the `idea_covidminder` slack channel
 - Contact Dr. Erickson to be added (if you haven’t already)

WHAT TO INCLUDE IN YOUR NOTEBOOK

- The code in your notebook should be well commented
- You should include thorough discussions of your methods and explanations of your results in the **markdown** sections (the non-code sections)
- You should always include enough detail for someone else to reproduce (and re-use!) your work!
- ****ALWAYS*** include useful titles and legends for your plots and tables!
- Nicely-formatted tables are always a bonus
- Your “default” for knitting notebooks should be HTML; this makes it easier to view and reduces errors.

EXPLANATION OF VISUALIZATIONS:

The goal of these visualizations is to examine nationwide disparities in COVID-19 factors having to do with risks, mediations (e.g. testing, hospital beds), and outcomes (e.g. deaths, cases). A common measure, the *disparity index* is used to represent the difference between the observed rate in the state and some baseline rate.

The advantage of the disparity index is that represents how far off a target standard the observed rate is.

Mathematically, $DI = \log(x/y)$ or $DI = \log(y/x)$ depending upon whether being above or below the target is preferred.

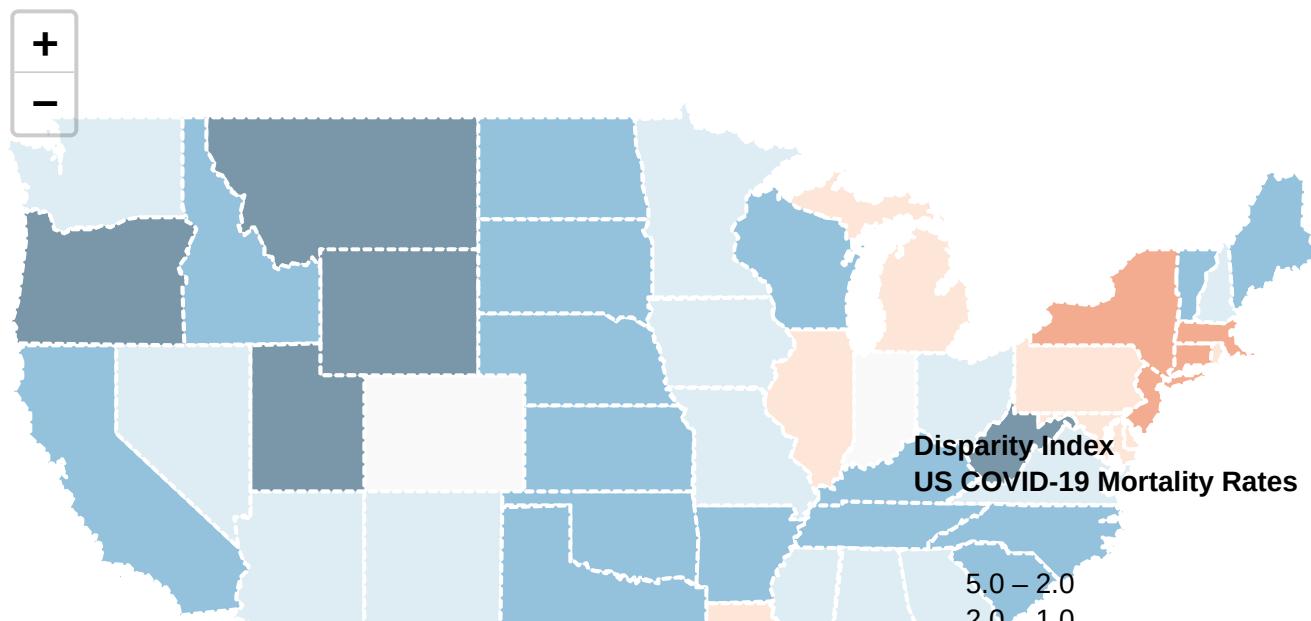
- In the case of hospital beds or rate of testing, x would be some state’s rate , and y would be the US rate or some rate we’re comparing against (e.g. South Korea’s testing or Italy’s hospital beds).
- In the case of mortality rates, x would be the target rate (e.g. some national rate, including the US), and y would be the individual state’s rate.

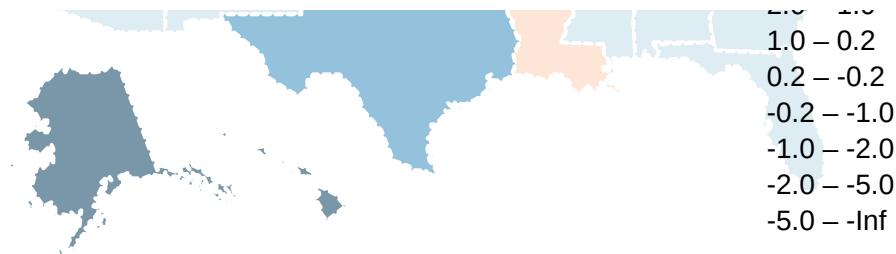
How do COVID-19 mortality rates compare across the United States?

This map compares the COVID-19 mortality rates of individual states with the US rate. This map is updated daily.

Here, **shades of red** indicate that a state’s COVID-19 mortality rate is higher than the US rate

Data source: JHU daily reports (<https://bit.ly/3dMWRP6>) (04-07-2020)





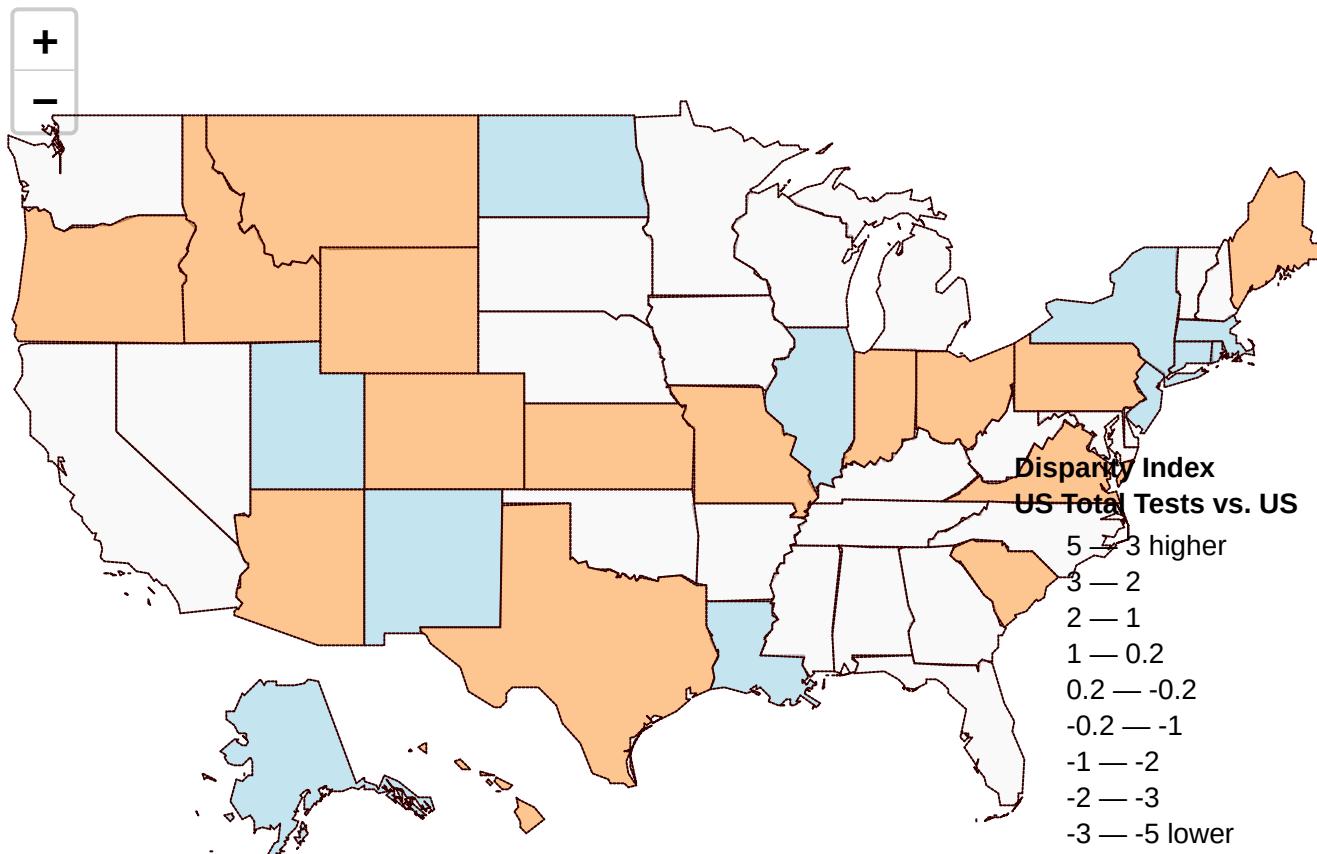
Leaflet (<http://leafletjs.com>) | Imagery from MapBox (<http://mapbox.com/about/maps/>) — Map data © OpenStreetMap (<http://www.openstreetmap.org/copyright>)

How do COVID-19 testing rates across the US compare with South Korea?

This map compares rates of COVID-19 testing in US states vs South Korea's testing rate. This map is updated daily.

Here, **shades of red** indicate that a state's testing rate is lower than the South Korean rate

Data source: The COVID Tracking Project daily reports (<https://covidtracking.com/api>) (04-07-2020)



Leaflet (<http://leafletjs.com>) | Imagery from MapBox (<http://mapbox.com/about/maps/>) — Map data © OpenStreetMap (<http://www.openstreetmap.org/copyright>)

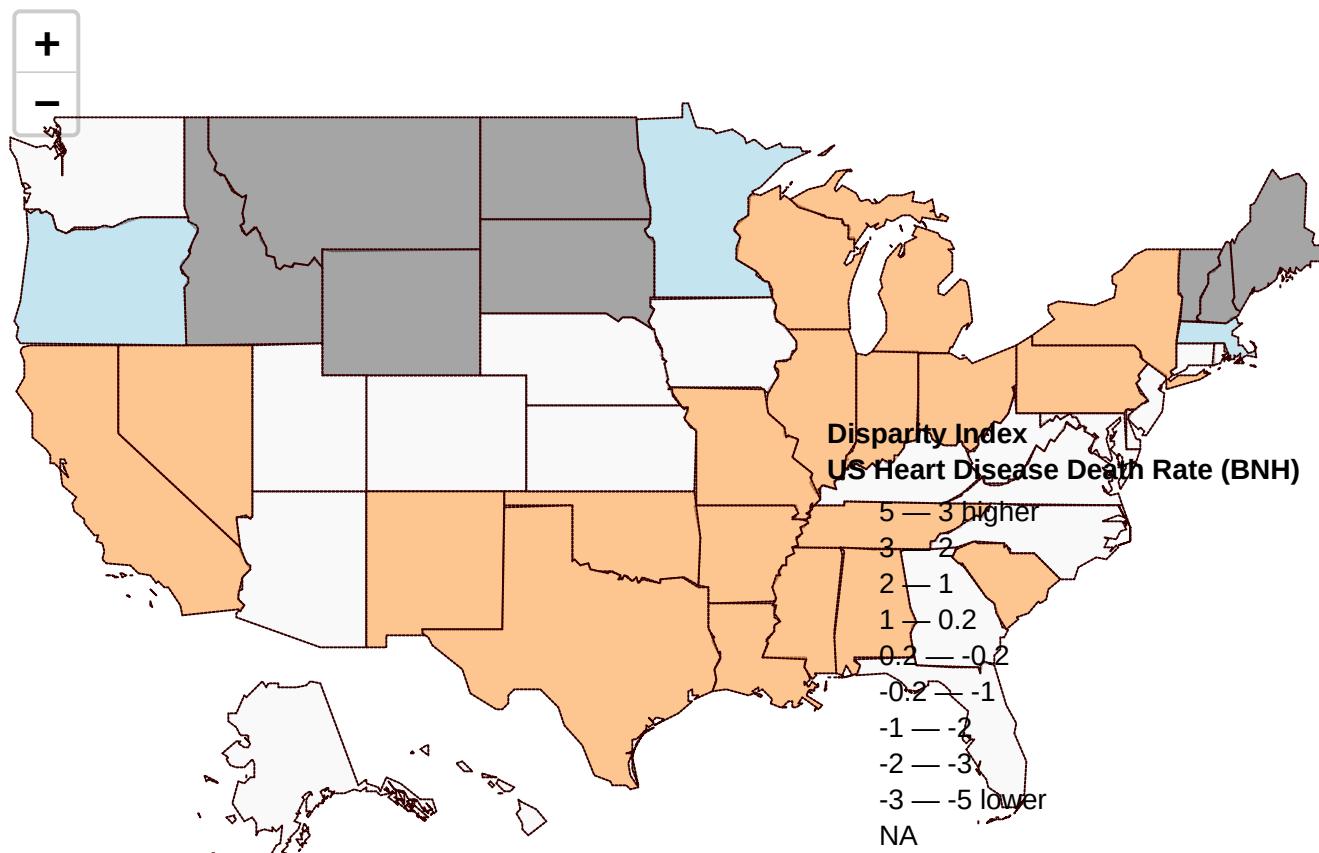
How do cardiovascular mortality rates across the US compare with the national average?

The map compares individual state mortality rates related to cardiovascular diseases (per 100k) with the US rate. In recent literature, COVID-19 risk has been linked to certain cardiovascular diseases, including hypertension. This map uses recent historical figures.

Here, **shades of red** indicate that a state's mortality rate from total cardiovascular diseases is **higher** than the US rate

Data source:

CDC (2017) (<https://bit.ly/2V1Zl3I>)



Leaflet (<http://leafletjs.com>) | Imagery from MapBox (<http://mapbox.com/about/maps/>) — Map data © OpenStreetMap (<http://www.openstreetmap.org/copyright>)

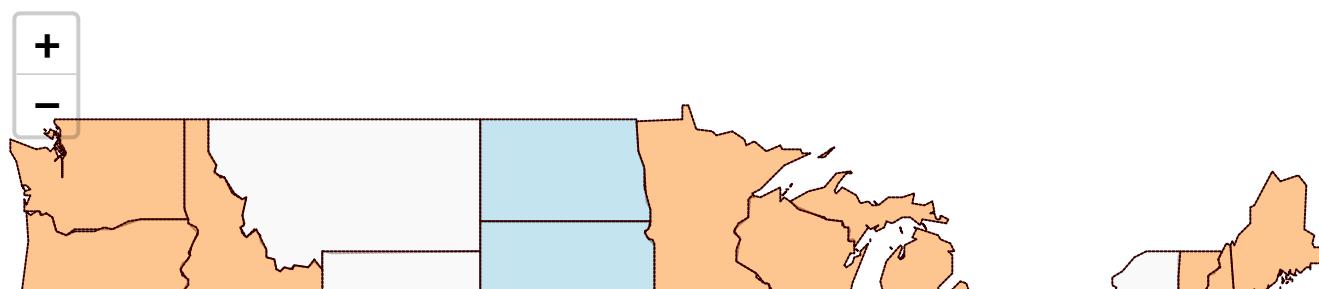
How does the availability of hospital beds across the United States compare with Italy?

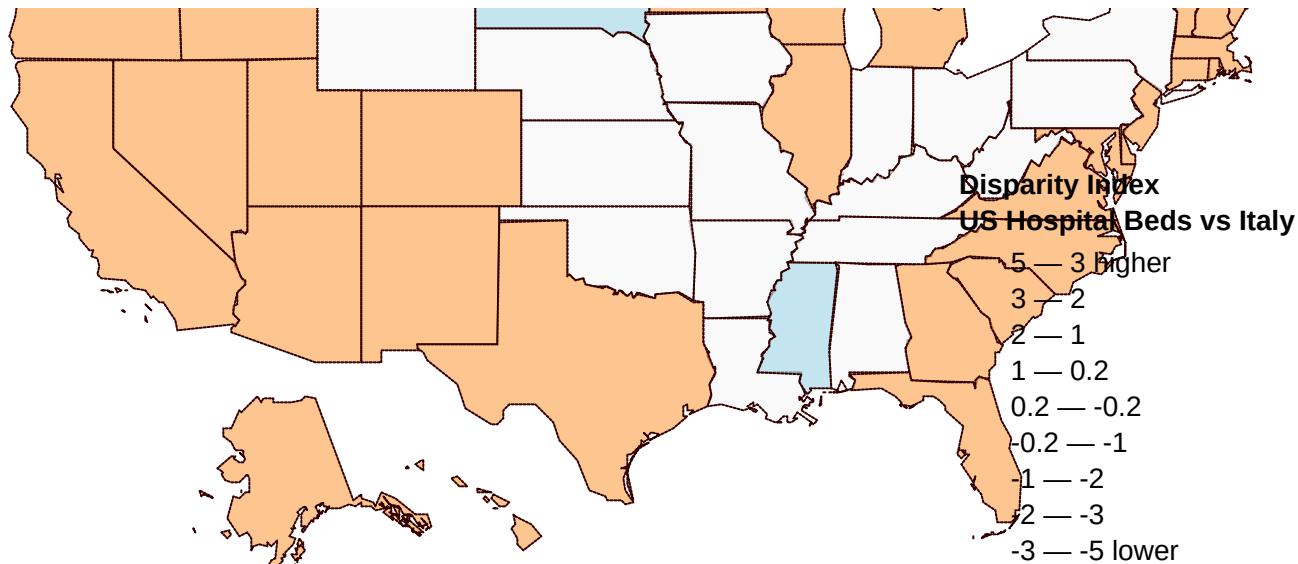
This map compares the availability of hospital beds in US states vs the rate in Italy (3.2 beds/1000). This map uses recent historical figures and does not reflect 'surge' capacity.

Here, **shades of red** indicate that a state's hospital bed availability is lower than the rate in **Italy**

Data sources:

Organisation for Economic Co-operation and Development (<https://data.oecd.org/healtheqt/hospital-beds.htm>) and Kaiser Family Foundation (<https://bit.ly/2V0CYLU>)





Leaflet (<http://leafletjs.com>) | Imagery from MapBox (<http://mapbox.com/about/maps/>) — Map data © OpenStreetMap (<http://www.openstreetmap.org/copyright>)

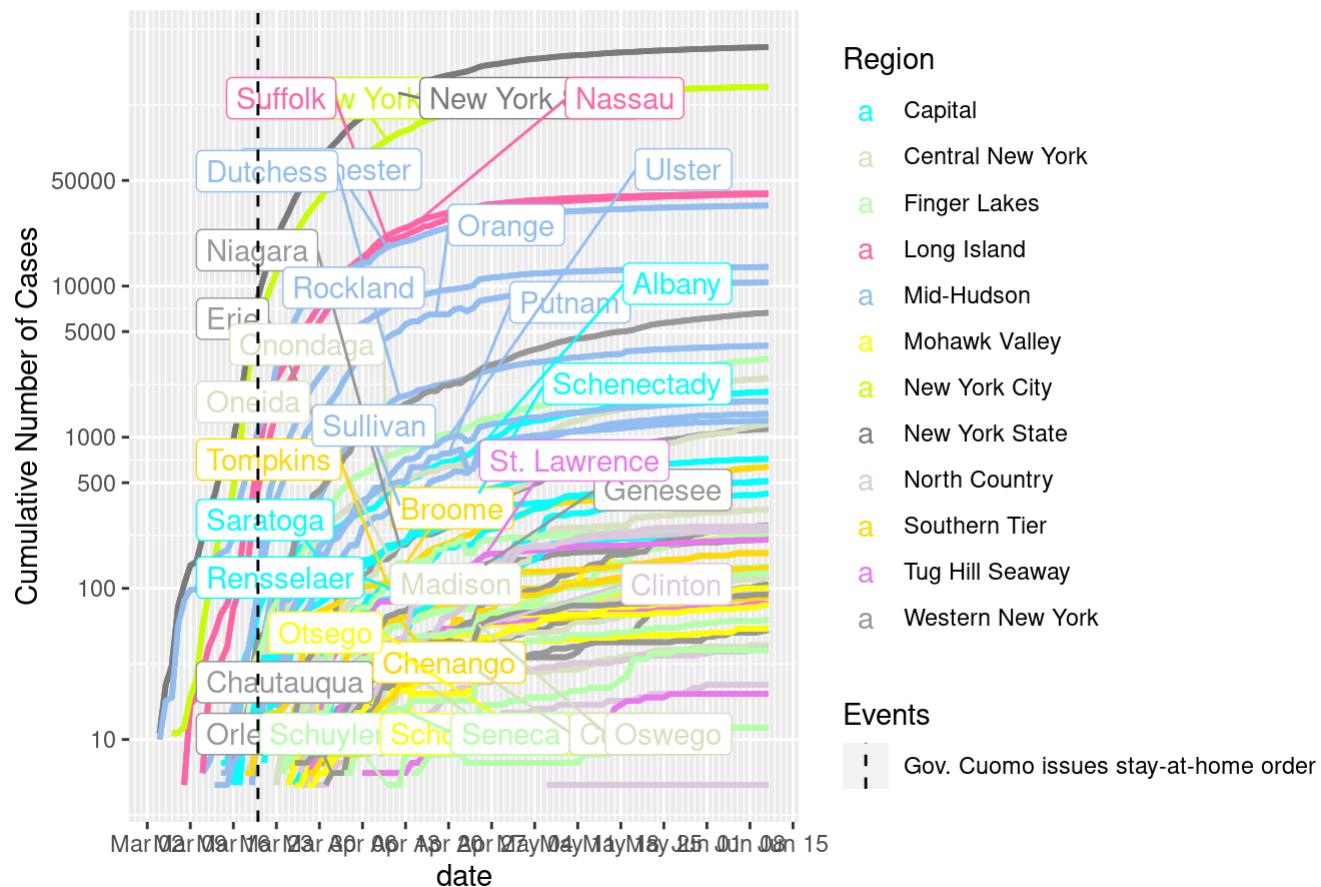
How have COVID-19 cases increased across New York State?

This plot shows the growth of COVID-19 cases across NY counties since early 2020. This data is updated daily.

Mouse over the plot to identify individual counties.

Data source: JHU daily reports (<https://bit.ly/3dMWRP6>) (04-06-2020)

New York State COVID-19 Cases per County (Mar-Apr 2020)



Accessing RStudio Server on the IDEA Cluster

The IDEA Cluster provides five compute nodes (4x 48 cores, 1x 80 cores, 1x storage server)

- RCS credentials (via registration in class); email John Erickson for problems `erickj4@rpi.edu`
- RStudio, Jupyter, MATLAB, GPUs (on two nodes); lots of storage and computes
- Access via RPI physical network or VPN only

RStudio GUI Access:

- `http://lp01.idea.rpi.edu/rstudio-ose/` (`http://lp01.idea.rpi.edu/rstudio-ose/`) or `http://lp01.idea.rpi.edu/rstudio-ose-3/` (`http://lp01.idea.rpi.edu/rstudio-ose-3/`) (RStudio)
- Linux terminal accessible from within RStudio “Terminal” or via ssh (below)

Shared Data on Cluster:

- All `idea_users` have access to shared storage via `/data` (“data” in your home directories)
- Shell access to nodes: You must access “landing pad” first, then compute node:
- `ssh your_rcs@lp01.idea.rpi.edu` For example: `ssh erickj4@lp01.idea.rpi.edu`
- Then, `ssh` to the desired compute node, e.g.: `ssh idea-node-02`