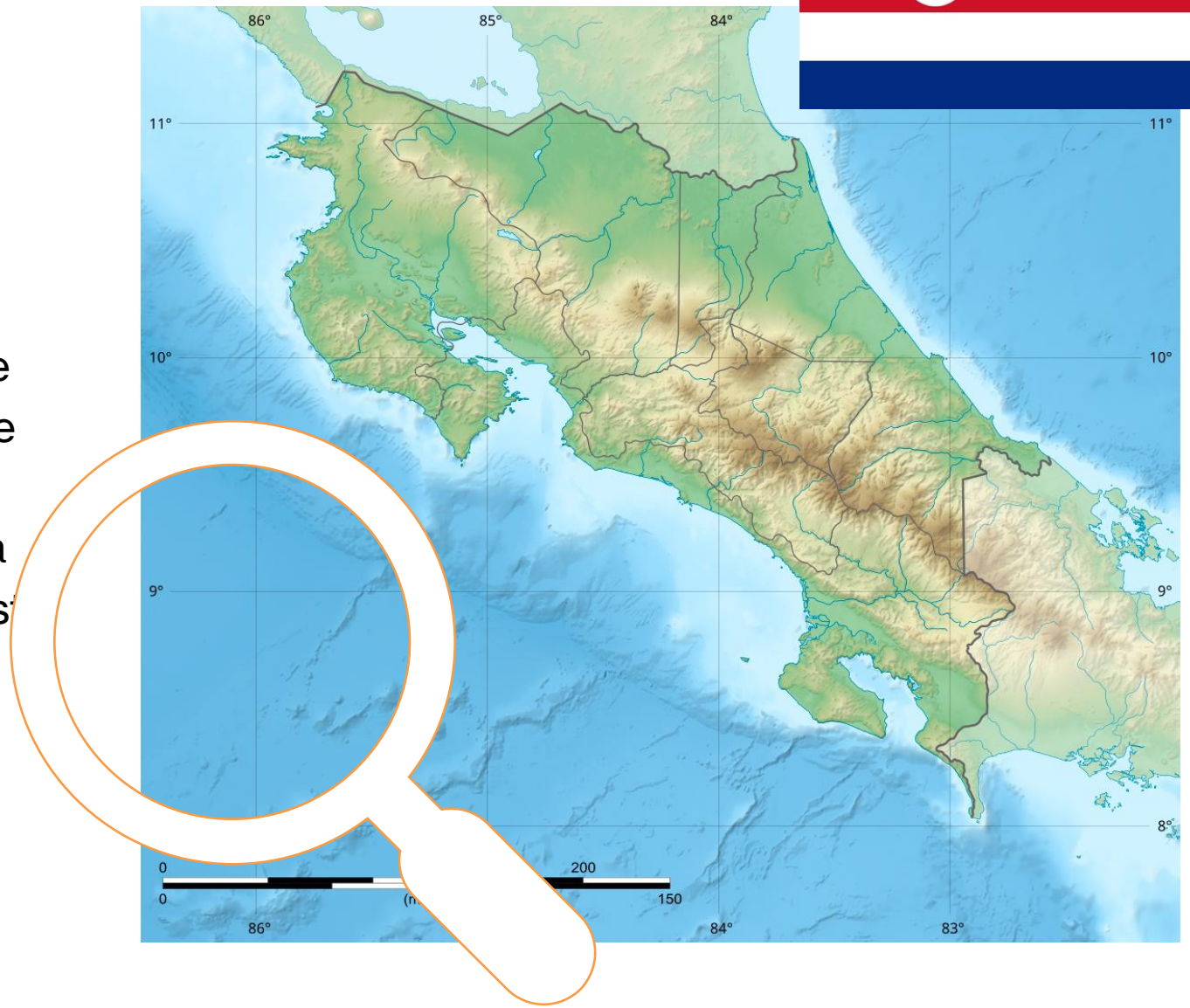
# Costa Rica Crime Data Visualization Web Application

Jenny Morgan

# What is it Exactly?

I created a data visualization web application encompassing an interactive choropleth map using government crime data from Costa Rica to visualize crime trends throughout each district in Costa Rica and total reported crimes from Post COVID (2021 – 2024) period.



# Why Did I Create This Project?

I was inspired to create this project when I moved to Costa Rica. I have family in Costa Rica who kept telling me how dangerous Costa Rica was and to be careful. They also told me to stay away from particular areas. My logical brain required substantial data to aid me to weigh their advice for myself. I also knew I may not be the only traveler of Costa Rica who would find this data useful in planning their trip while in Costa Rica, and / or simply to conceptualize the amount of crime in each district within Costa Rica.

# My Biggest Blocks During the Project

While creating my project, I had no idea the pipeline of creating a web application, I only knew that I wanted to create a data visualization map that could help people. Particularly travelers in Costa Rica.

One of my biggest problems was attempting to successfully mend my local project to the host network- I first tried Render and later with Railway.app. When connecting my project to Render I persistently received errors since the port could not be found. I learned this was because if the project does not render fully by a certain amount of time (Flask renders the whole project at once and then returns the compiled project) and since my project has lots of data heavy files, Render assumes my program is not working and kills the program. This is due to Render's free tier having strict CPU and Memory limitations. I transferred my project to Railway.app and I have not had any issues since then.

Merging the crime data successfully was a challenge due to the fact I had to merge the GeoJSON file (the file which contains all of the map coordinates for the districts) with the crime data files, but because some districts in Costa Rica are not unique, I had to merge the data from the District and Canton columns (which is similar to a county). Data can be very sensitive to being merged so I learned to be exact when merging and to check edge cases, like District names being changed.

# My Biggest Successes During the Project

MY WHOLE PROJECT!

1. The interactive choropleth map itself is something I am proud of creating
2. The loading feature was also fun
3. The data extraction part was super cool. I love how numbers can always paint a story. This is one of the reasons I was so drawn to only getting data from Post-COVID. Everything post-COVID has changed, the effects from the pandemic has rippled throughout countries and societies and can be seen even in the microcosm society of Costa Rica.

# The Biggest Problems & How Tackled them

NULL DATA!

Some data did not merge properly and resulted in null data for those particular districts. I investigated what lead to this issue and found these unique reasons:

1. Some districts changed the name of their Canton in the past decade, therefore the Canton name on the GeoJSON did not match up with the crime data in the police reports. For these cases, I simply changed the name

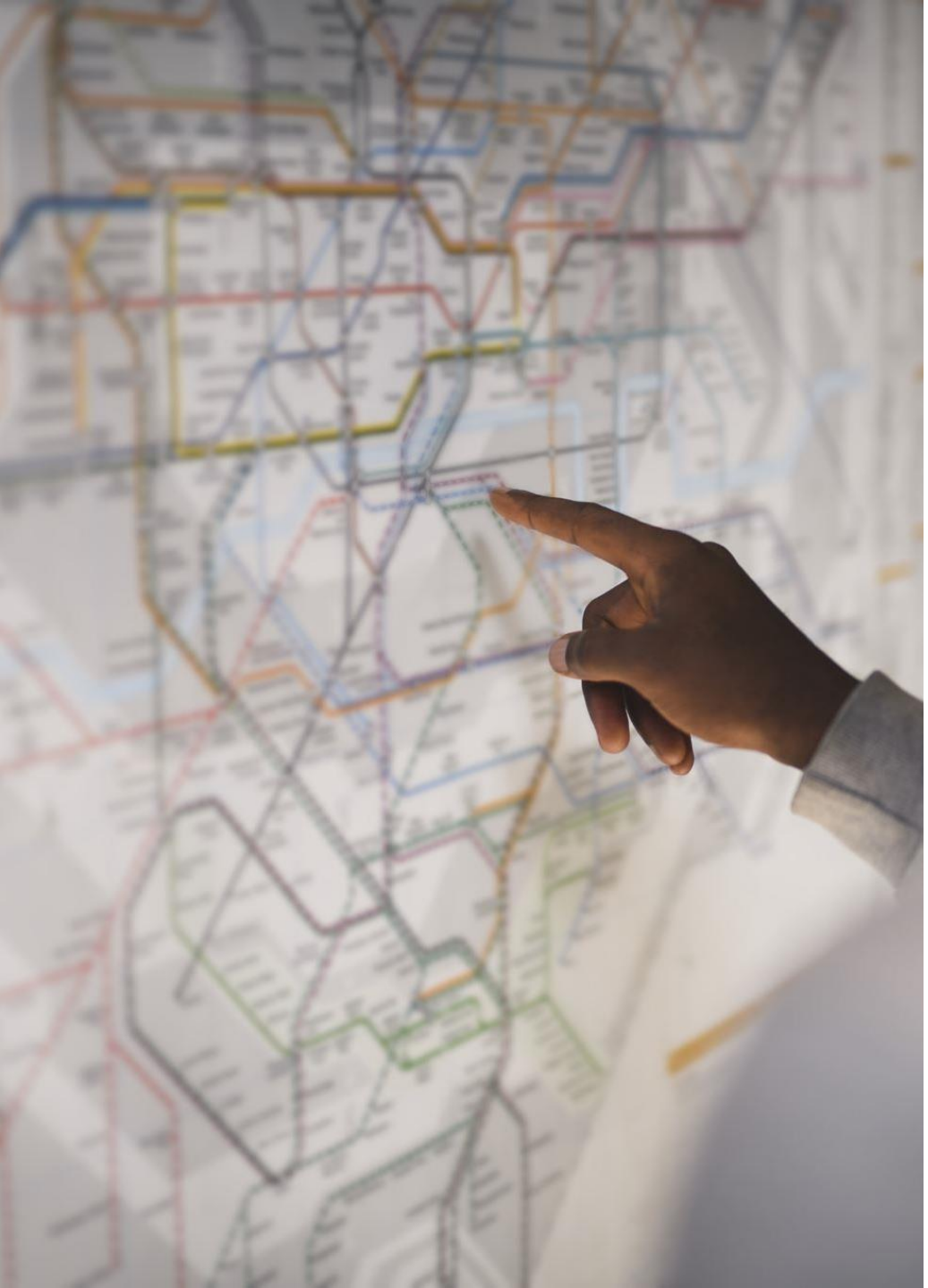
- One example is the Aguirre canton, they changed their legal name to Quepos in 2015
- Another example is Valverde Vega canton which changed their legal name to Sarchi in 2019
- Rio Cuarto canton was created in 2019

2. There were discrepancies in spellings between Cantons in government crime data vs the GeoJSON. For these cases, I changed the GeoJSON name to match the crime data

- For example, Vazquez de Coronado is the spelling in the GeoJSON and legal spelling but the spelling in the government crime data is Vásquez de Coronado, which could mean there is a discrepancy between how the locals spell the name and the official spellings. Or simply human error.
- Mercedes District in Guacimo canton was a similar story . The crime data spelled the district as Merecedes, rather than Mercedes which is the official government legal name for the district.

3. Clear human error

- Sierpe district is in the canton of Osa, but was mistakenly labeled to have the canton of Puntarenas in the GeoJSON.
- Angeles district of San Rafael was misspelled in the GeoJSON and is actually legally named Los Angeles district of San Rafael canton



# The Biggest Problems & How I Tackled them

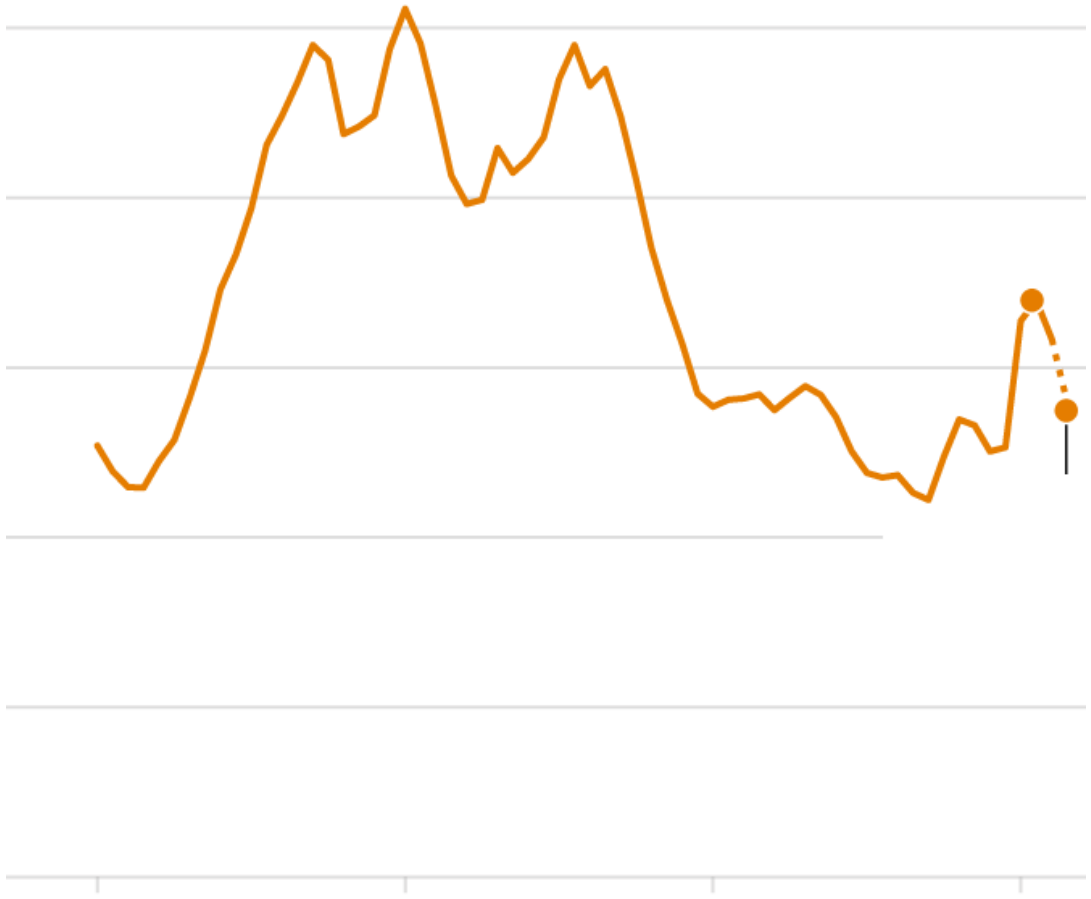
DATA MERGING ISSUES!

I had issues merging certain district data when the data was present in the crime datasets and the GeoJSON. I learned that the pd.merge() function automatically resorts to its default settings of merging how="inner" which merges each dataset based on columns which are present in each, since some districts have null data for certain years (meaning, no reported crimes!), the district gets lost after the merge, incurring in null data for the whole district.

The fix: how="outer" for the merge of each years overall total crimes.

# What Functionality I Would Add Next

I would calculate the the crime rate for each district and add it to the popup functionality. The crime rate for each district would also be useful to know.




I learned so much from this project. Inclusive, how numbers paint a story, and the technical skills acquired through the pipeline of creating an interactive map based on real world data and launching a web application
Folium is a great library to use to create maps, incredibly helpful and they have all the tools I needed. It was all about reading the documentations and learning the technical parts of their functions
When a free hosting network is not working, try another!
There are a few districts in Costa Rica which either changed names, or changed Canton names in the past decade
Large scale data visualization is impossible to get accurately, but we sure can come as close as possible!
We must always ponder human error and edge cases dependent on the project mission within data sets.