

Farhan Alavi April 2nd 2024

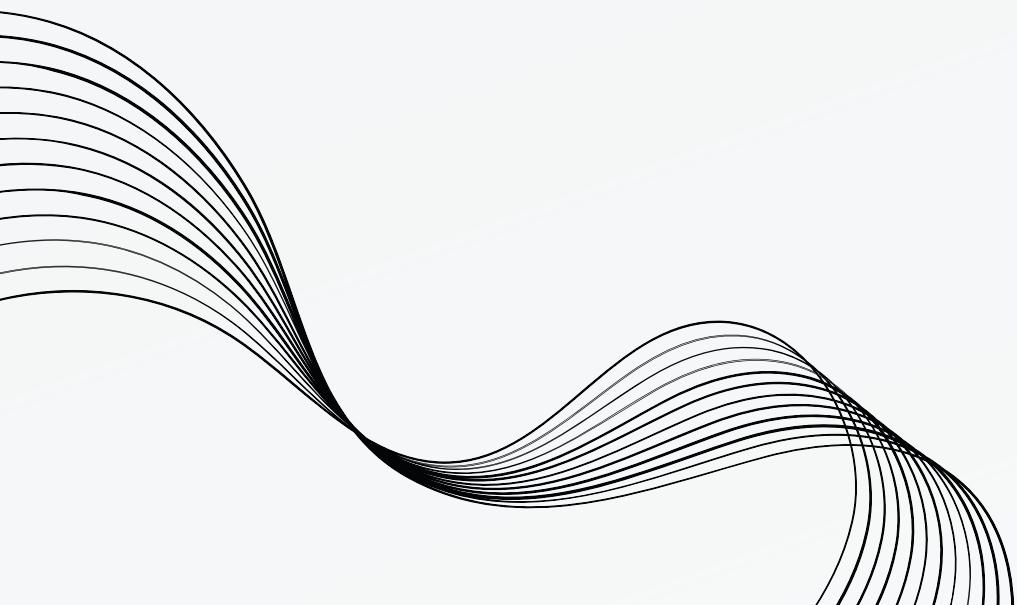
Infectious Disease Cases in California from 2001 to 2022

INTRODUCTION

There were **511,874** cases reported in California in the last **21 years**.

What is the Data Science approach?

- What is the problem?
- How will we solve it?



WHO BENEFITS FROM THIS?



Can utilize a predictive Machine Learning model to forecast future cases and use it as a study tool.



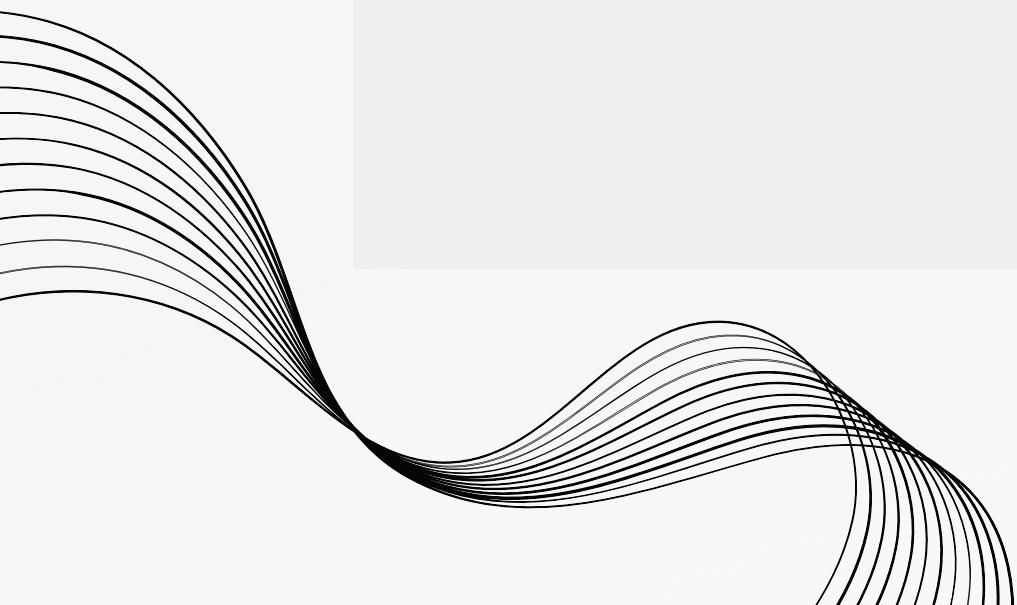
Visuals can help non-medical professionals to follow along as well.

INITIAL DATA FINDINGS

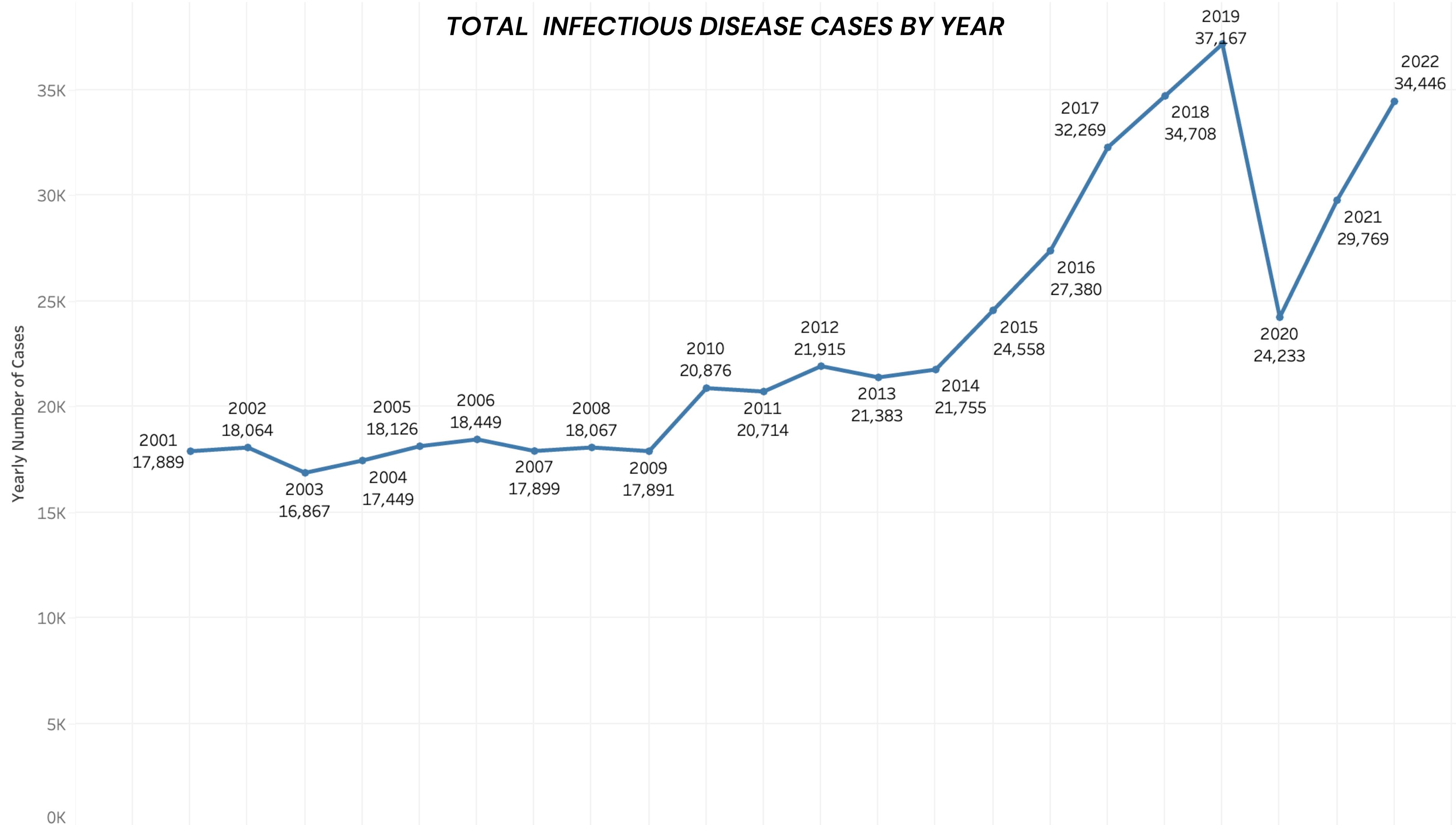
Before cleaning - Initial size: **180,000**
rows x **6** columns.

Post-cleaning- Final size: **62234** rows x
78 columns.

Problem areas while cleaning?



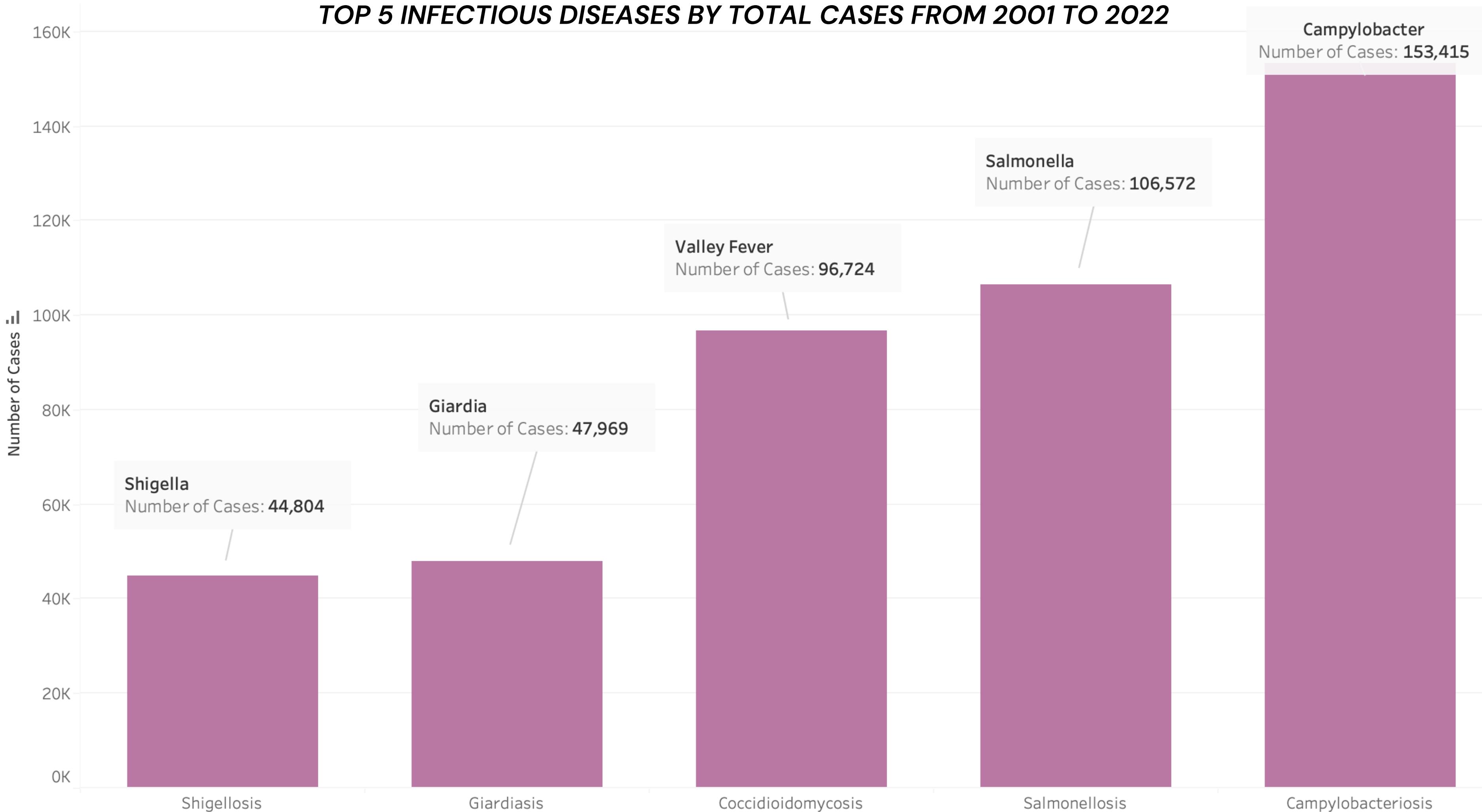
TOTAL INFECTIOUS DISEASE CASES BY YEAR



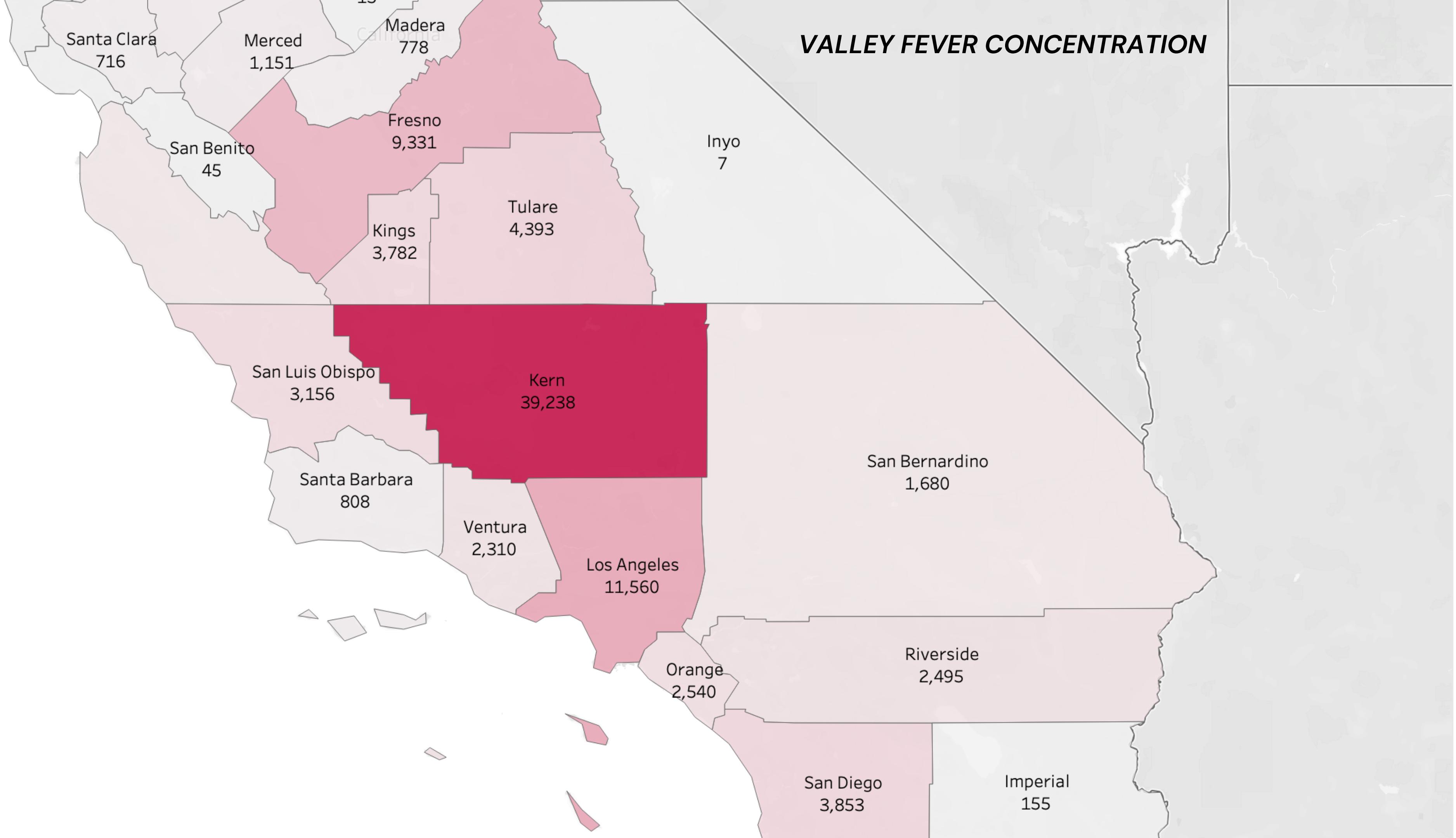
CASE DISTRIBUTION BY COUNTY

107,192 Los Angeles	28,210 Orange	25,976 Santa Clara	22,801 Alameda	22,147 San Francisco	21,498 Fresno
47,984 Kern	16,772 Riverside	12,255 San Mateo	8,045 Stanislaus	6,529 Santa Barbara	6,502 Sonoma
45,832 San Diego	14,479 San Bernardino	11,348 San Joaquin	5,693 Monterey	4,154 Merced	3,788 Placer
	13,994 Sacramento	10,452 Ventura	4,919 Kings	3,103 Imperial	2,077 Napa
	13,811 Contra Costa	10,435 Tulare	4,707 Solano	2,941 Butte	1,465
			4,465 Marin	2,810 Madera	896
				2,580 Yolo	825

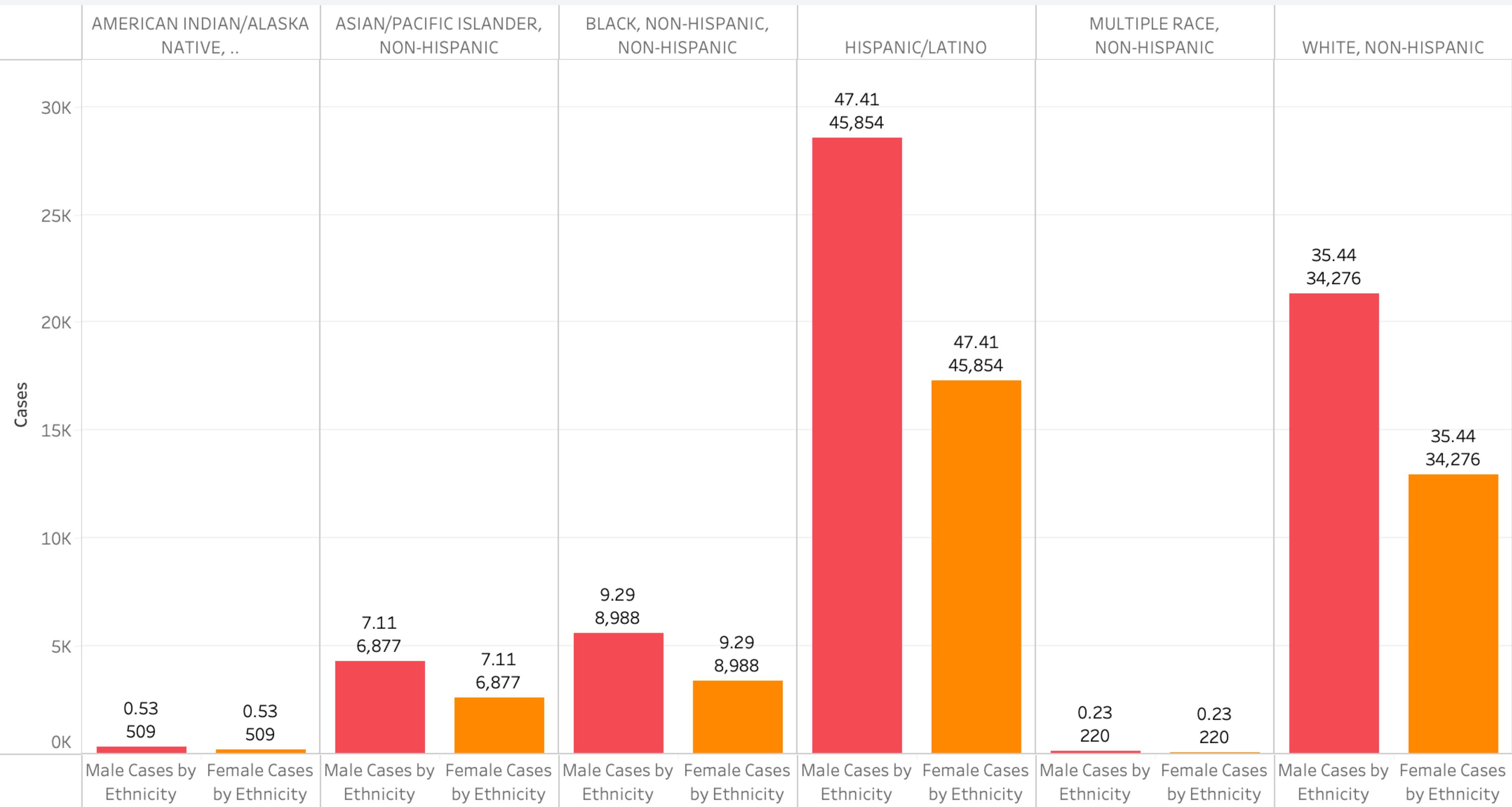
TOP 5 INFECTIOUS DISEASES BY TOTAL CASES FROM 2001 TO 2022



VALLEY FEVER CONCENTRATION



VALLEY FEVER ETHNIC BREAKDOWN



NEXT STEPS

- Back to the drawing board.
- Polish the baseline models and boost accuracy.
- See how close to the real 2023 data I can get without exposing my model to it.

**Thanks
for watching!
Questions?**

