## **QGIS Documentation**

* Pulled shapefiles from [California Open Data for the county and state boundaries](https://data.ca.gov/dataset/ca-geographic-boundaries/resource/b0007416-a325-4777-9295-368ea6b710e6?inner_span=True) and [data.gov for the PUMAs](https://catalog.data.gov/dataset/2020-cartographic-boundary-file-shp-2020-public-use-microdata-areas-for-california-1-500000). Later pulled a [somewhat crude open source clipped-to-shoreline version](https://www.arcgis.com/home/item.html?id=2f227372477d4cddadc0cd0b002ec657) – one did not appear to be available online.
* Set QGIS CRS to 4269 (based on NAD 83, which seemed to be a consistent choice for Northern California mapping)
* Uploaded the shp files into QGIS as layers
* I exported a csv with stats for the northern california counties in 2019 and 2022, and the change in these stats, from R ("ncal\_counties\_2019\_2022.csv")
* I had to do a lot of manual cleaning of the data in QGIS – it was reading almost every single field type in the csv as the wrong type of data
* Joined the csv to the counties map layer 3 times to create 3 choropleth maps.
  + then, used graduated symbology to show % change in:
    - super commutes (2019 vs 2022) - based on point of origin (home address)
    - super commutes (2019 vs 2022) - based on point of destination (work address)
    - total population (2019 vs 2022)
  + Applied labels by rule-based labeling to highlight counties of interest (outliers in one way or another) (had some trouble with centering labels inside of counties but adjusted within QGIS to mostly fix.
* Export maps via print layout, where I added legends, etc.
* Use Adobe Express to crop blank space from the maps.