Week5 Activity-Visualizations

Qingyuan Li

2022-09-25

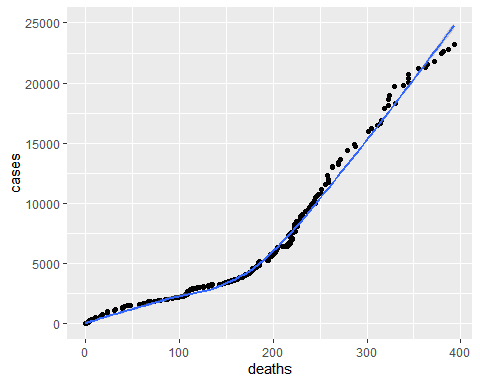
Section 1: Description of Data

The dataset I am using is the 2020 year’s documented daily COVID cases and deaths data collected and published by New York Times. The primary data published are the daily cumulative number of cases and deaths reported in each county and state across the U.S. since the beginning of the pandemic, and I picked only the 2020 csv file. The data is the product of dozens of journalists working across several time zones to monitor news conferences, analyze data releases and seek clarification from public officials on how they categorize cases.This data is very helpful to study the most impactful COVID-19 cases/deaths surges and also the geographic time series of how it spreaded all over the US. The format is csv (a flat file). It is delimited by comma. It is not binary.

Section 2: Reading the Data into R

Section 3: Clean the Data

Section 4: Visualization I - Washington county’s cases and deaths

 In Washington County, the COVID-19 case number grows faster than deaths at first, but become a bit slower in the long run.

Section 5: Visualization II - The distribution of COVID-19 cases by the end of 2020 at county level

The distribution is very right skewed. Most counties are having a below-average number of cases.

