

Owid-covid-2020-visual-analysis

It is significant to be seen that there are six attributes in the world wide covid data monthly in 2020. Respectively location, month, case fatality rate, total cases, new cases total deaths and new deaths. The location column covers the most of countries on the world, this means that the covid is a virus has strong infectiousness in order to be spread in the worldwide. The raw data demonstrates that for the most of locations, not every month bring cases or death. The second plot is made through changing the x-axis, which converted the total cases of origin data to the log scale because of an order of magnitude difference.

Both plots are made up by fatality rate of cases(y-axis) and total new cases(x-axis) data by locations in 2020 from owid-covid-data-2020-monthly csv file.

For the scatter-a-plot), it shows that the majority of points are located on a grid, which is $0-2 \times 10^6$ at x-axis and $0-0.05$ at y-axis. This implies that most of locations not only have a relatively few new cases amount, which is under 2×10^6 , but also have a lower fatality rate, which is under 5 percent. Besides, it can be seen that the locations have numerous cases basically exist very low fatality rate. This means the covid doesn't have high lethality rate in general. Oppositely, some few cases locations have a relative high death rate in 2020. A fact can be inferred from this plot is no location maintains a mass of cases with a superhigh fatality rate.

For the scatter-b-plot, it can be seen very clearly that the scatter-b has a difference with the scatter-a although they exhibit the same content. However, a very intuitive information can be obtained from the plot is that for the total number of new cases in 2020 of locations between 7.5-unit and 15.0-unit, occupied a large proportion of the whole number of locations with a under 5 percent fatality rate. This means that many locations in the world did have the new cases from covid but only few death cases occurred, which has the same research results as the first plot. Besides, there is another group of characteristic crossing plot on the plot that represents there have been no deaths despite owning the new cases in some locations. Moreover, It can be seen that there are totally only six locations, which have a more than five percent fatality rate. This result is more in line with the real situation.

Furthermore, some unregular points is showed on the both plots. For example, both plots appear a point has a quiet high fatality rate location. This may represent the level of medical system in the location is relative lower than the basic standard in the world, so that the medical workers are not able to mitigate or treat the damage caused by covid virus.

There is an obvious difference between them. The first plot's points are basically squeezed into a 1×1 grid and a great deal of points are lied on or nearby the 0-value x-axis. In this case, it is difficult to demonstrate the accurate information the task may look for so that the difference or relationship between each point may not be discovered, only a rough conclusion may be received. Nevertheless, points on the second plots are equably distributed on the plot by contrast. Using the log-scale for the total new cases values are extremely narrow the difference in order of magnitude. As a result, the scatter-b is easier for readers to search a precise relationship between the fatality rate and total new cases values for each location.

Two plots in Part A Task 2:

