List of variables

1. Age
2. Age Group
3. Date Symptoms Started
4. Date Report was confirmed
5. Date they died
6. If they were hospitalized
7. Barangay
8. City
9. Province
10. Region

What do I want out of this report?

I want to focus on the lethality of covid in the Philippines. Collecting data related to covid is a very complicated matter. Things such as vaccination rates, vaccine types, availability of testing kits, reporting of confirmed cases, measures taken by local government, wearing masks, social distancing, washing of hands all play a part in how we measure data related to covid. I myself got covid for the first time in April this year, and when I tested positive, while I didn’t report it, I did quarantine in my room at home testing every few days until my results were negative. From my experience alone, I suspect that there is an underreporting of covid cases (source?), especially now with how things have opened up compared to the past, as we continue with this new normal. There are less restrictions in place now in comparison to the start of the pandemic, with establishments allowing full capacity, less enforcing of face masks and face shields, and the ease of domestic and overseas travel. So, when we look at data of the confirmed cases we can take this to be a minimum of the real number of cases. But this doesn’t mean that we can’t make use of the data available to us. If we assume that age doesn’t have a large impact on someone’s choice to not report their case of covid, then we can assume that the number of cases across the different age groups would be proportional to their real value. And so, with the data we have, making comparisons between age groups and covid would still be a reasonable action to take. Similarly, if we assumes that within NCR the city you live in plays little part in your likelihood to report your case of covid, then we can make comparisons between the cities in NCR too. However, one aspect that I feel would be difficult to work around would be people reporting their cases of covid less over time. So comparing certain aspects of covid by different time periods can be challenging. Except for one thing. The death rates. I believe (source?) that the information we have related to people dying due to covid is close to accurate. When people die, and they’re brought in for autopsy to determine the cause of death covid can be found, especially in more recent times, since we’ve grown more knowledgeable on the matter. So if we assume that the number of deaths related to covid is close to their actual value, and we assume that there is an underreporting of covid cases, the lethality we find is a maximum. And so we can find something closer to the worst case. Additionally, if our assumption that the likelihood that someone would report their case of covid is decreasing over time, then that means at our current point in time, the lethality of covid should be at its peak, but according to our data the lethality is at the lowest it has ever been. Meaning that the lethality is truly quite small when considering the underreporting.

I want to show that over time, covid has become less lethal. Since we have data on the admit rate to hospitals due to covid, we can also see how often covid hospitalizes people, and like with lethality, this is a maximum possible value. So we can have an idea of the percentage of people dying and getting admitted to the hospital due to covid. So with all that said, this analysis seeks to find trends between the following

1. Age and Getting Covid
2. Age and getting hospitalized
3. Age and fatality

Side discussions

1. Date Rep Conf vs Date Onset
2. Time delay between Date Rep Conf and Date Onset
3. Time delay between Date Rep Conf/Date Onset and Date Died
4. How graphs show that the number of cases & number of deaths are proportional
   1. Assuming that the number of deaths is accurate
   2. Assuming that the number of cases is truly proportional to the number of deaths
   3. This suggests that the underreporting is close to consistent
      1. If the underreporting was inconsistent there’d be more noise in the graphs
   4. Cases Recorded = Covid Cases \* Underreporting Ratio
   5. Number of deaths = Covid Cases \* Lethality