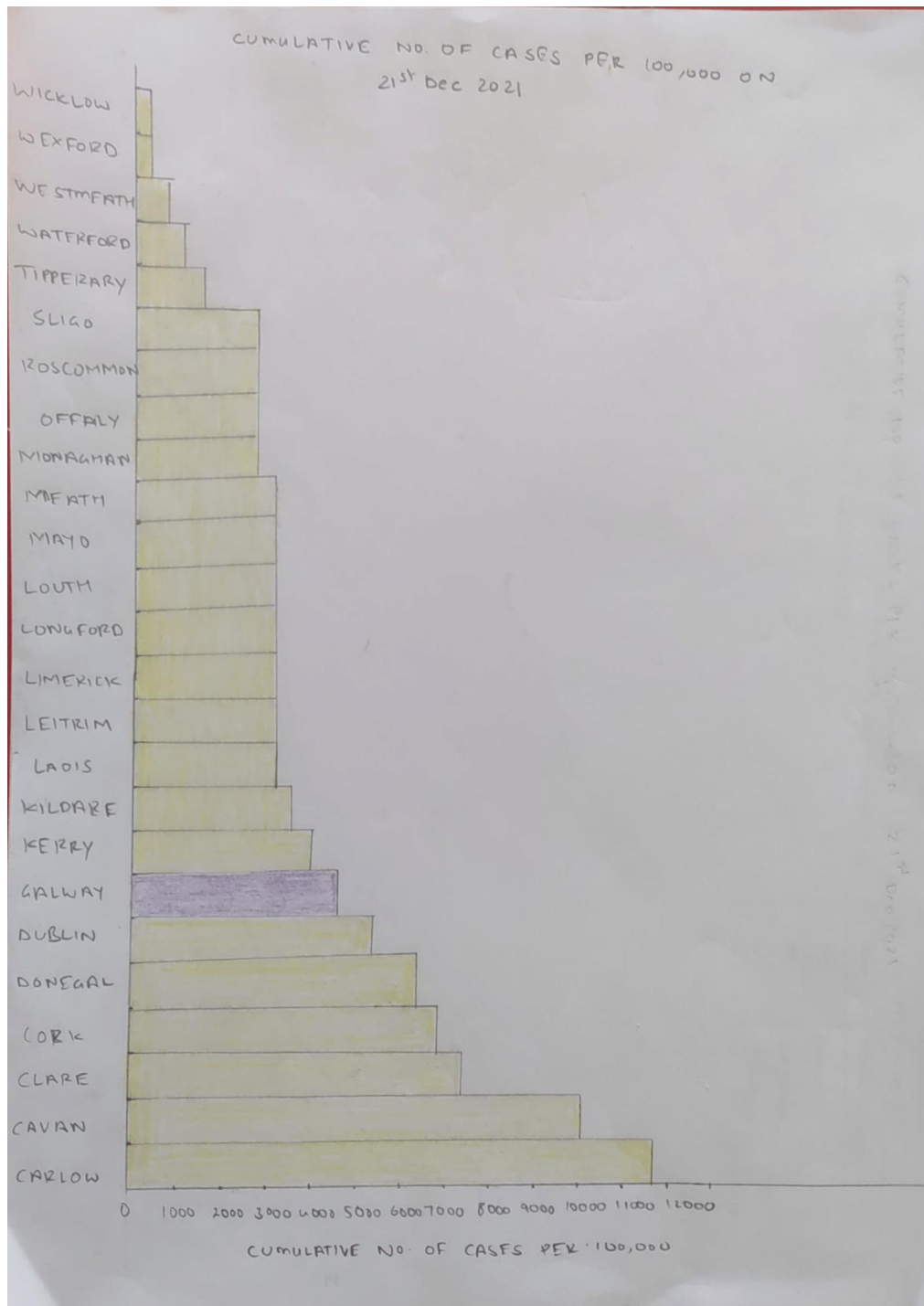


Assignment 2

1. A visualisation that allows the reader to accurately compare the cumulative number of cases per 100,000** of population per county on 21 December 2021. County Galway should be highlighted.

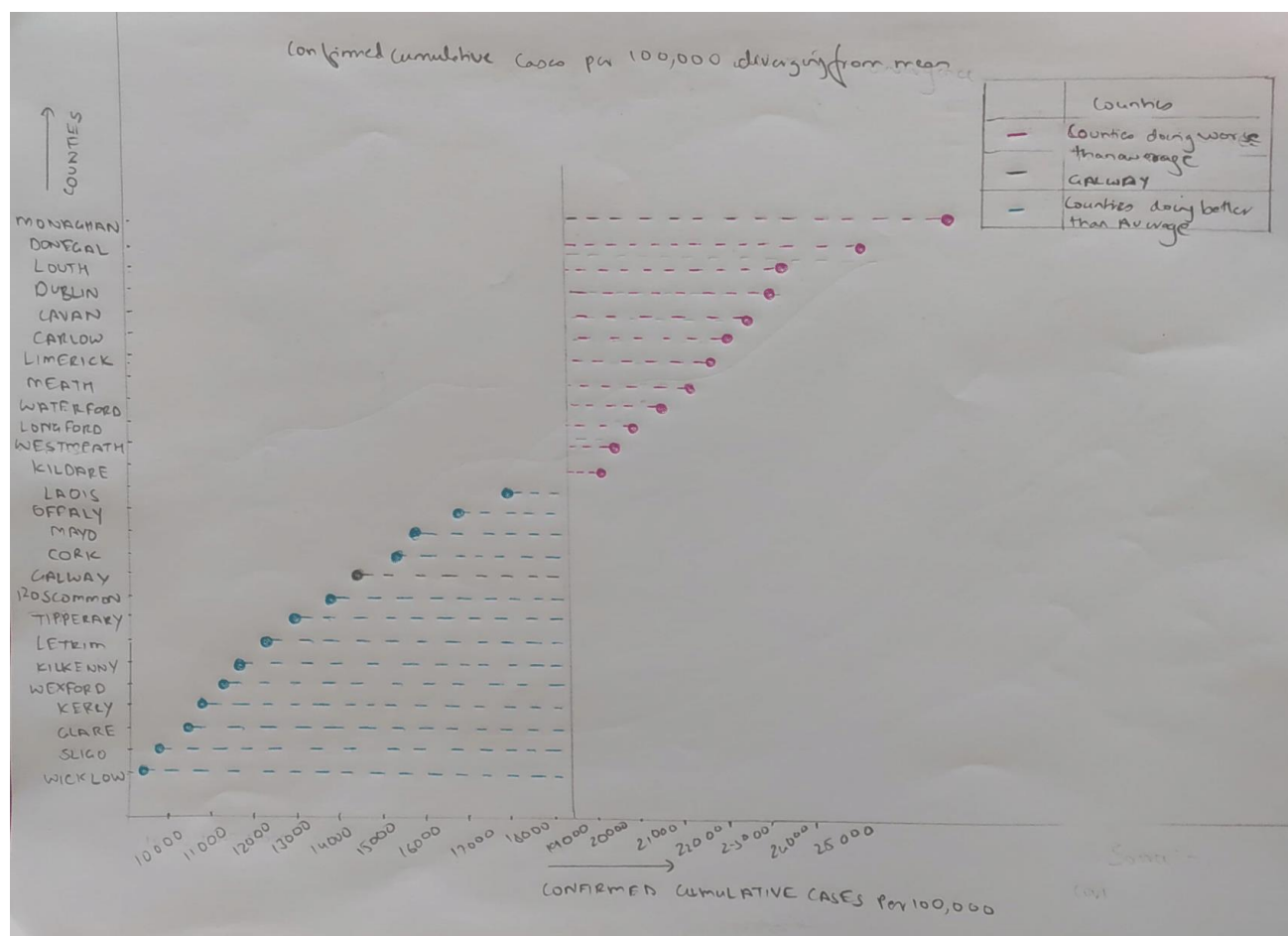
Answer: A simple bar plot with the bar for county Galway highlighted should work for this problem. As the data points to be plotted are for a single day, we can plot a bar plot sorted in order of the cumulative number of cases per 100,000 so that it is easy for the eye to follow the order with Galway highlighted.



Calculations Required: Data for date '21-12-2021' should be filtered and sorted by the value in column 'ConfirmedC_per_100k'

2. A visualisation that allows the reader to read how each county diverges from the mean cumulative number of cases (per 100,000) in the country as of 21 December 2021. You may also use a daily figure in this section. County Galway should be highlighted.

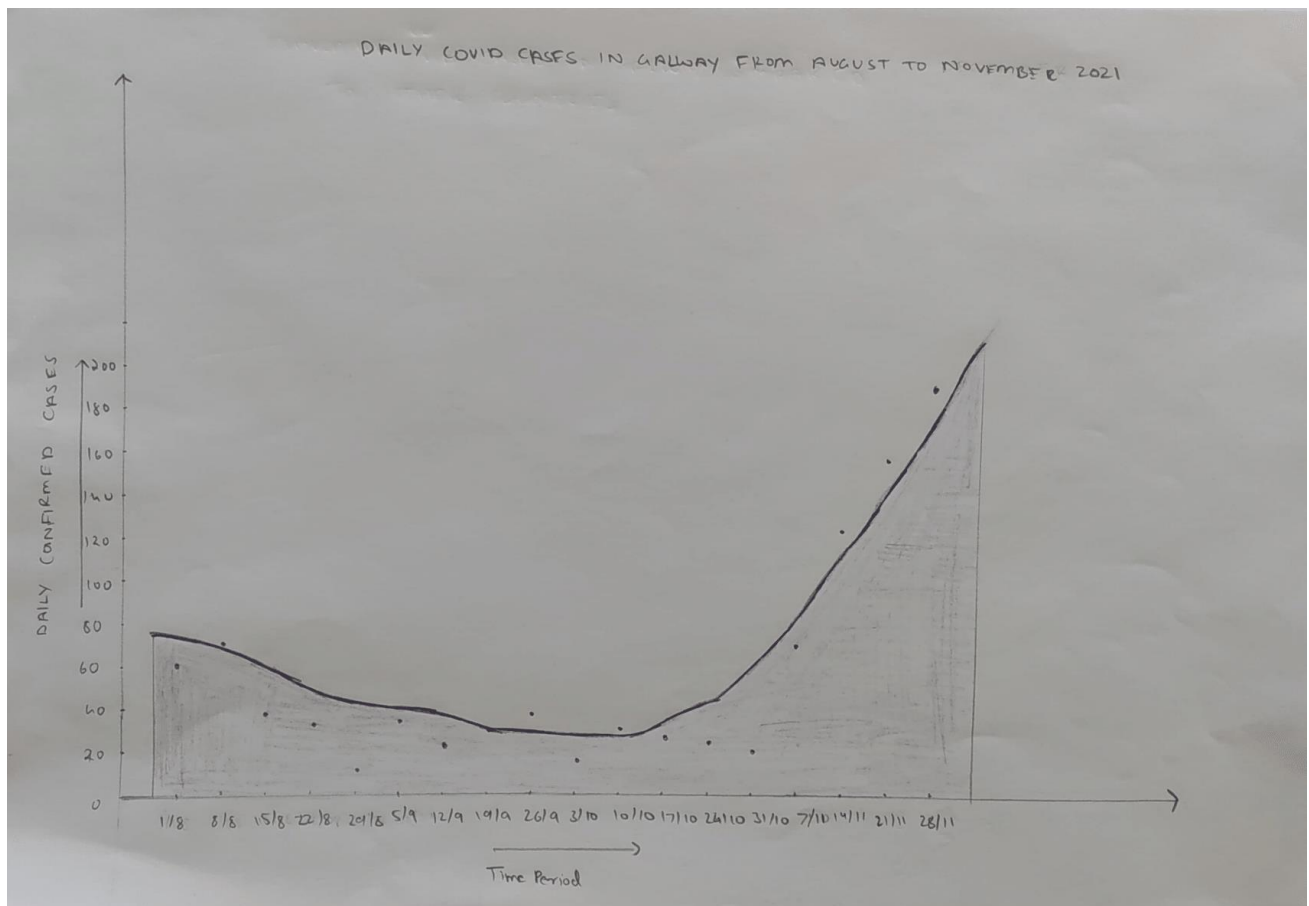
Answer: As we have to show how each county diverges from the average no. of cases in Ireland, we can use a diverging bar chart to showcase how much each county deviates from the mean value. A diverging chart usually has a color encoding for points above the mean and points below the mean. Hence, different colors would be used for counties for their divergence from the mean. Also, the County Galway would be highlighted in a different color.



Calculations Required: Same as previous code.

3. A visualisation showing the daily number of confirmed covid cases in one county in Ireland for 18 weeks. This visualisation should help the reader to perceive the trend in the data.

Answer: A trend line can be created using the `geom_smooth()` function with a filled area under the line. For time-series data, a trend line with the dots can show the trajectory of information changes with respect to time effectively. As the data showcased is cumulative i.e summed over time highlighting the area would give the idea of change in the volume of cases with respect to time.

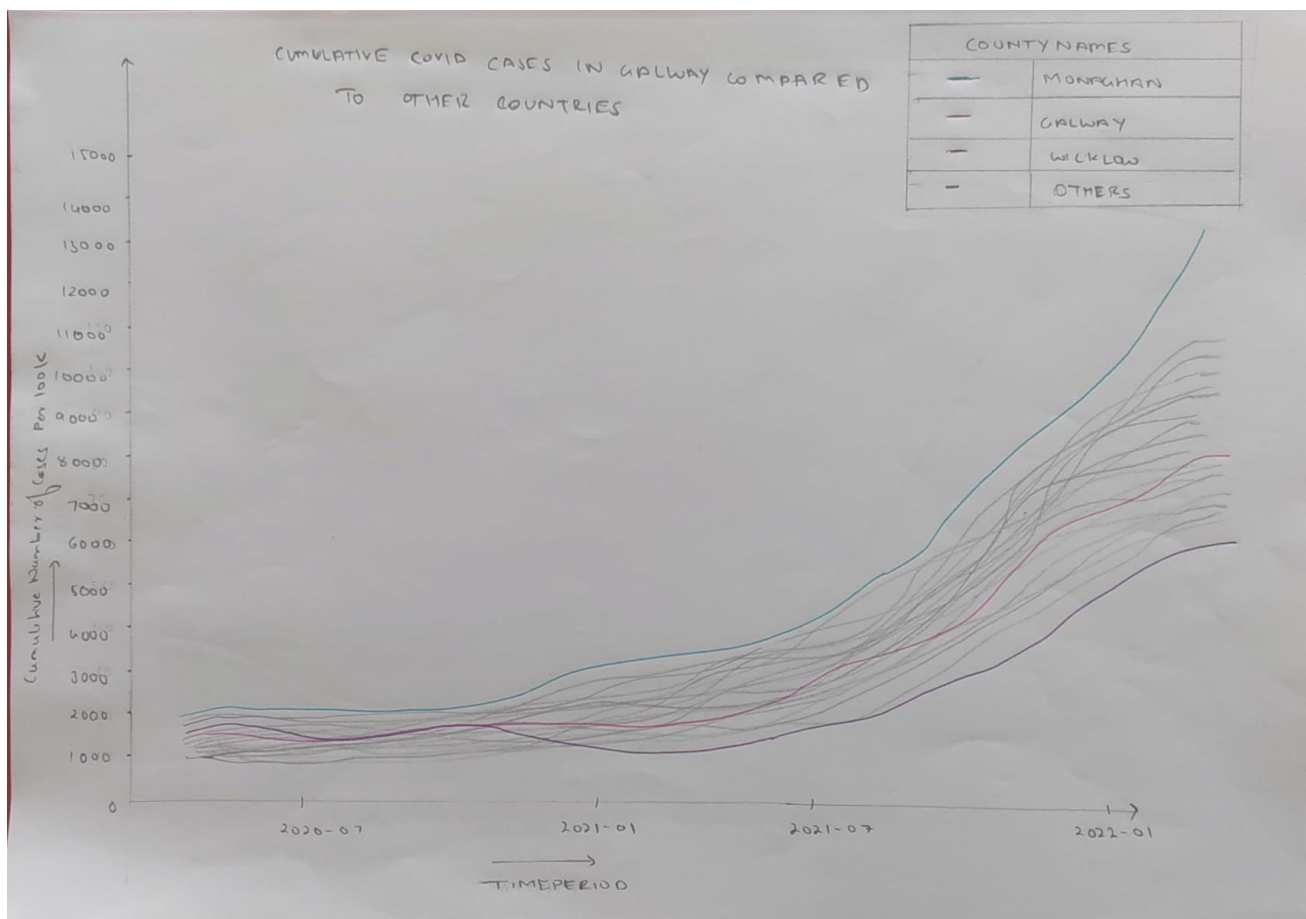


Calculation Required: One of the Counties data (Galway in this case) should be filtered along with a time period of 18 weeks.

- A visualisation that highlights the cumulative number of cases per 100,000 in Galway and two other counties representing counties that have had the lowest and highest number of cases per 100,000 over the full timeline of the dataset. The visualisation must also show the cumulative case number for all other counties in Ireland in the same plot. However, the three selected counties (Galway and two other counties) must be highlighted)

Answer:

As per the problem statement, we have to plot trend of cumulative number of cases per 100,000 with three counties highlighted. We can plot trend lines for all the counties with respect to confirmed cases. However, it would look messy and data would be indistinguishable for all the counties because there are a lot of them. We can highlight the three counties and faded the remaining ones for better visualisation.



Calculation Required: We need to find the counties with highest and lowest cumulative cases, we can do that by finding the county with minimum and maximum values in cumulative column.

5. A choropleth visualisation of the counties of Ireland showing total new confirmed cases (per 100,000) for a 4-week period (of your choice) for each county. The choropleth should show how each county diverges from the mean number of new confirmed cases (per 100,000) per county for that 4-week period.

Answer: To show how the number of new confirmed cases diverge over a period of 4 weeks we need to find the difference between cases occurring between that period, and then plot in on a choropleth how the difference in number of cases diverges from the mean number of cases for all the counties. As the goal here is to show the number of cases of a disease, a diverging palette like 'RdYlBu' palette for RColorBrewer library would be appropriate as the dark colors help in conveying the spread of disease and cool colors to show a decrease in cases. Counties where the number of cases have decreased during the 4-week period would be cooler in color compared to counties where the number of cases have increased.

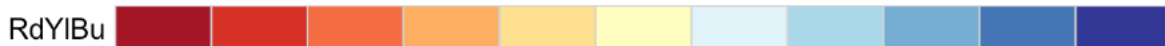


Image Source: Data visualisation, week 6 lab CVDv2.1.html

Calculations required: We need to calculate number of new cases along a period of 4 weeks and then plot the deviation from the mean of the values for all counties. The 4 week period for plotting would be from 2021-01-01 to 2021-01-29