

Contents

1 Step 3 Look at the datasets	1
1.1 Region	1

1 Step 3 Look at the datasets

1.1 Region

As we can see on the website, Region metrics are available for regions of England. I am interested in the South West and metrics that start with “New”:

areaCode
areaName
areaType
date
newPeopleVaccinatedFirstDoseByVaccinationDate
newPeopleVaccinatedSecondDoseByVaccinationDate
newPeopleVaccinatedThirdInjectionByVaccinationDate

We have additional columns. Let’s look at them.

- For **areaCode** unique value is *E12000009*,
- for **areaName** unique value is *SouthWest*,
- for **areaType** unique value is *region*.

So, we do not need to look at them in the future because these columns are used for filtering that we have already done on the website.

Let’s prepare data for the plotting.

- Rename columns and columns
- Add the column MonthYear
- Create long table

date	MonthYear	Dose	Count
2022-01-26	1.2022	First	986
2022-01-25	1.2022	First	899
2022-01-24	1.2022	First	723
2022-01-23	1.2022	First	1035
2022-01-22	1.2022	First	1822
2022-01-21	1.2022	First	1085

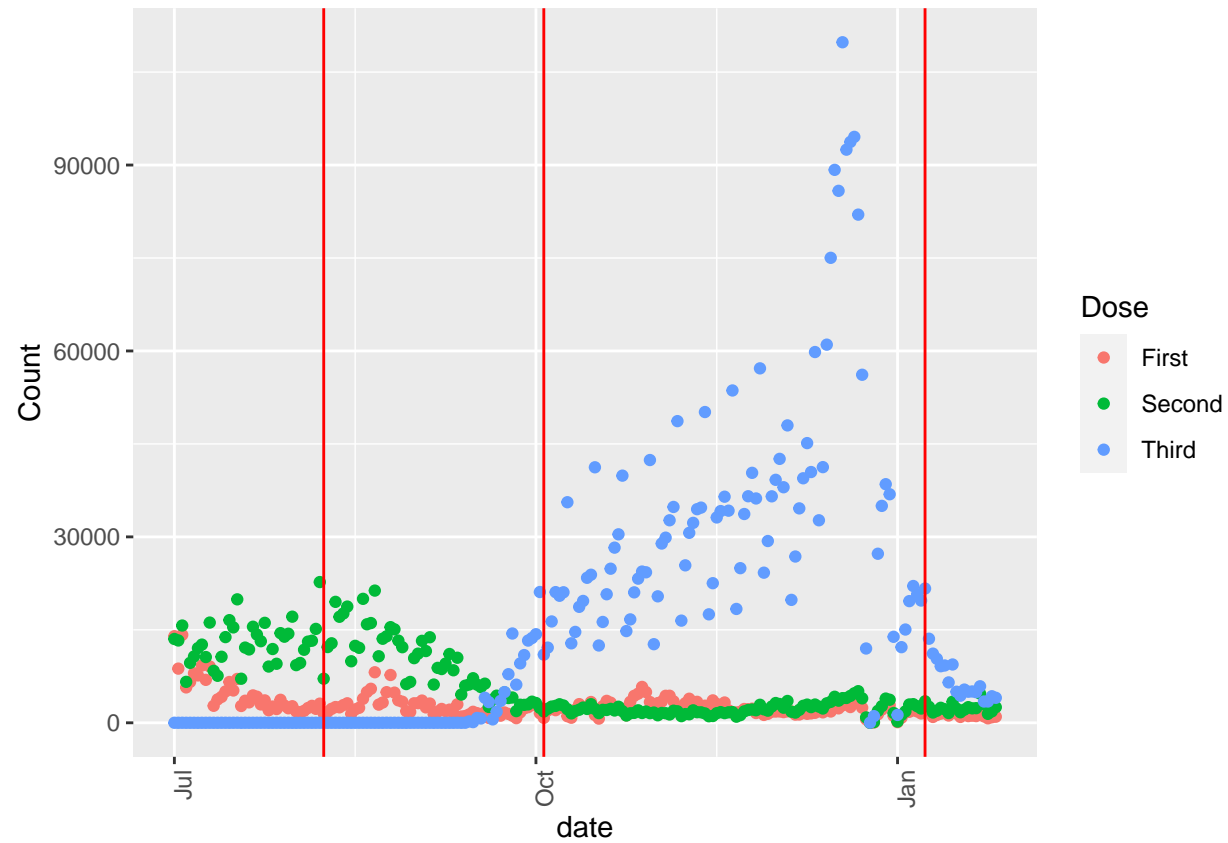
Let's plot something.

```
myFirst <- as.Date("2021-08-08", format="%Y-%m-%d")
mySecond <- as.Date("2021-10-03", format="%Y-%m-%d")
myThird <- as.Date("2022-01-08", format="%Y-%m-%d")

region_set_long_period <- filter(region_set_long, date >= as.Date("2021-07-01", format = "%Y-%m-%d"))
```

1.1.1 Question 0

Figure 1: My caption

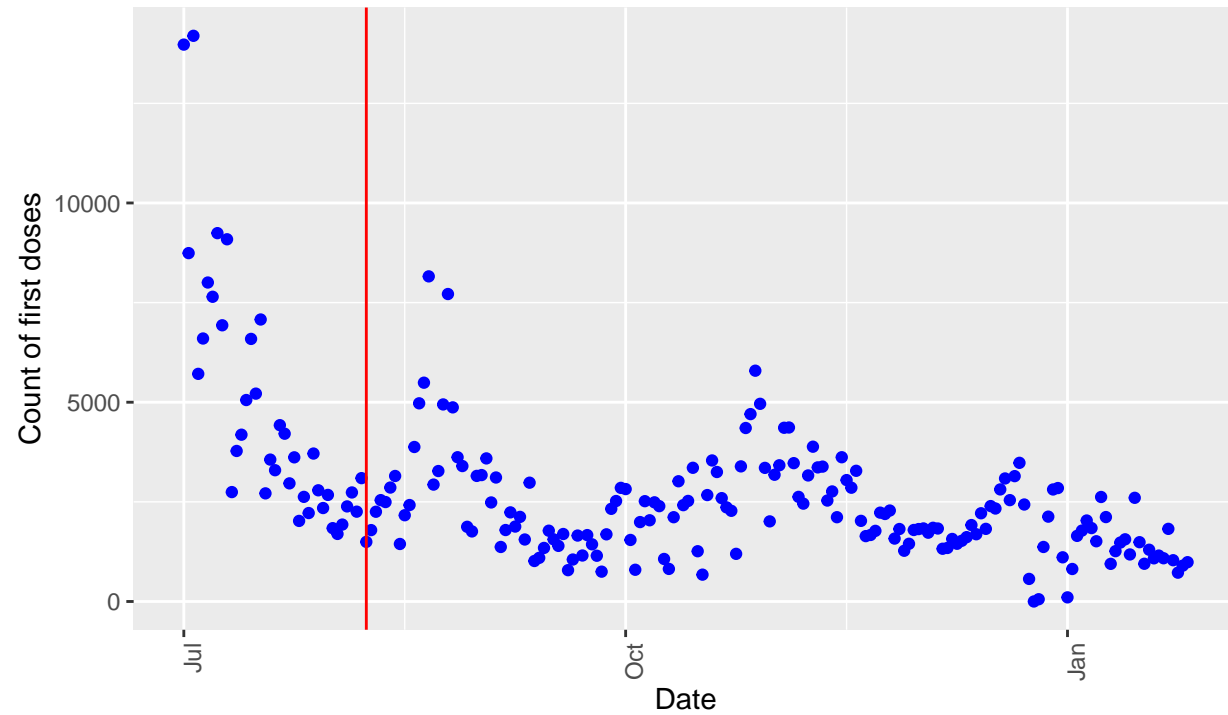


The result is not beautiful because of the active growth of the third jabs count at the end of 2021.

Let's plot them separately.

Vaccination in South West

The first dose

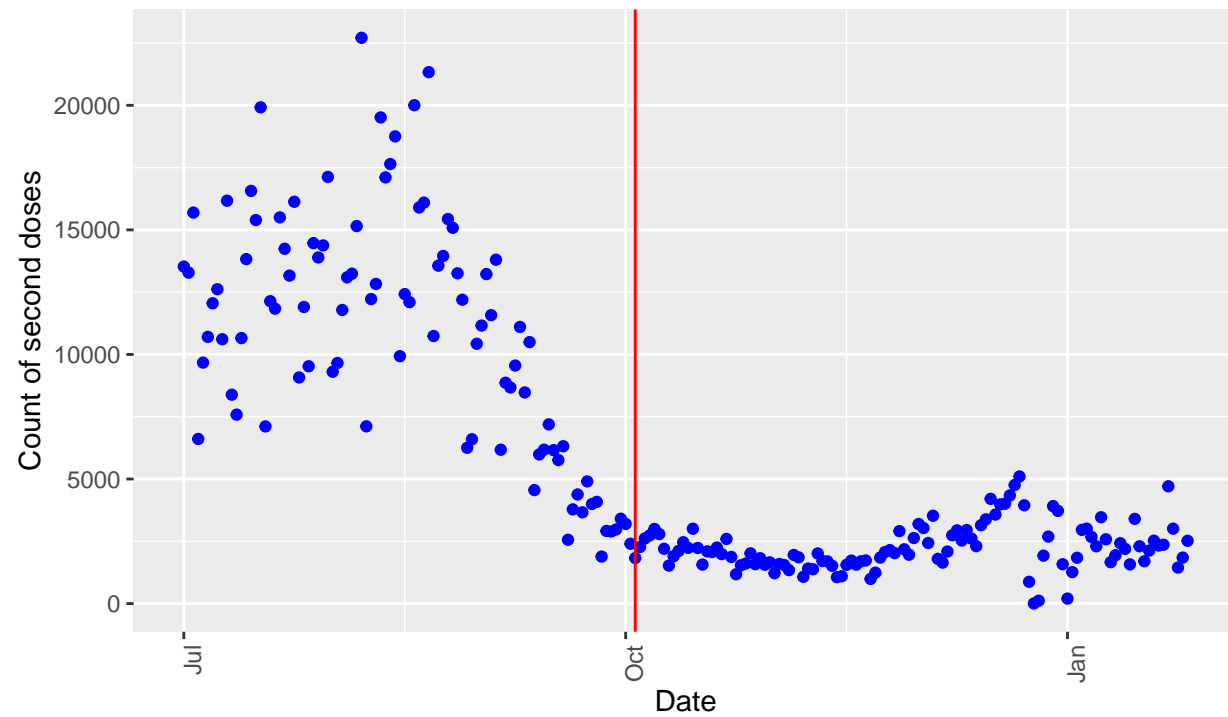


More information <https://coronavirus.data.gov.uk/details/about-data>

It is so interesting why the graph is wavy. 1496 people got their first jabs with me.

Vaccination in South West

The second dose

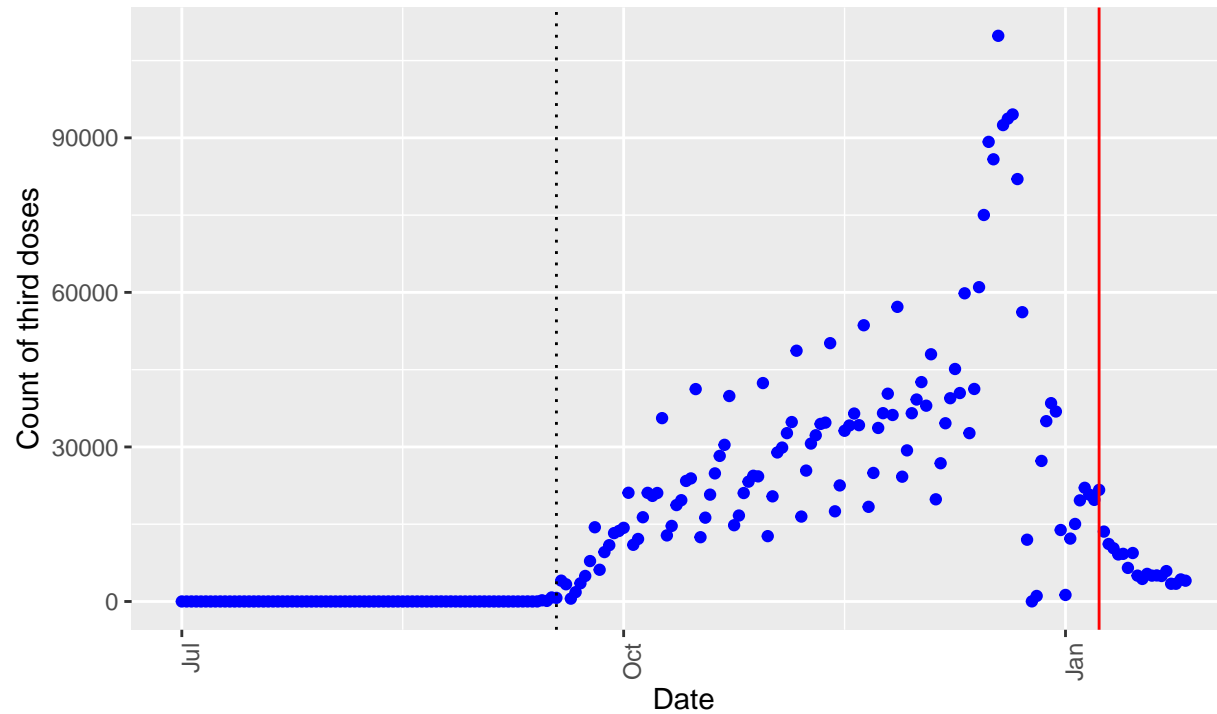


More information <https://coronavirus.data.gov.uk/details/about-data>

1828 people got their second jabs with me.

Vaccination in South West

The third dose



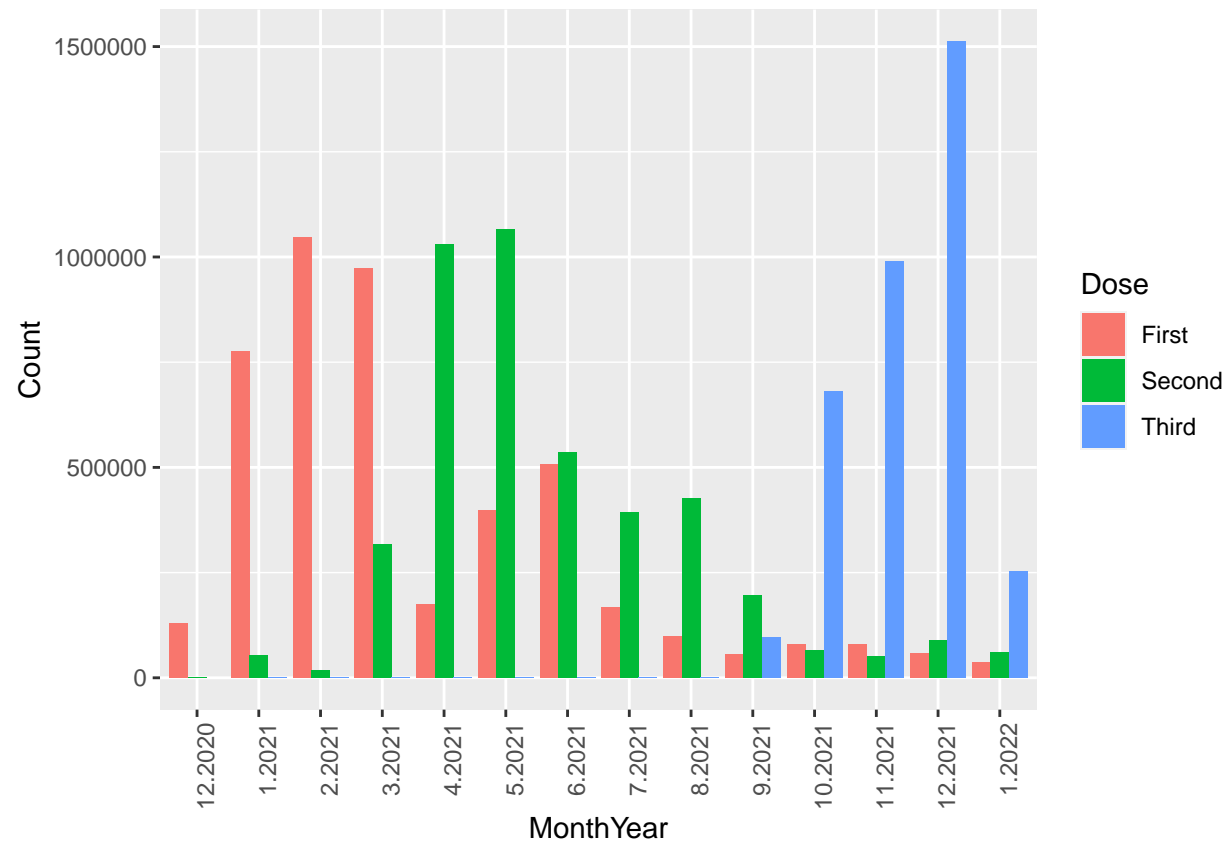
More information <https://coronavirus.data.gov.uk/details/about-data>

21664 people got their third jabs with me. We can see when the active phase of vaccination by the third dose started.

Let's calculate the date. ("2021-09-17")

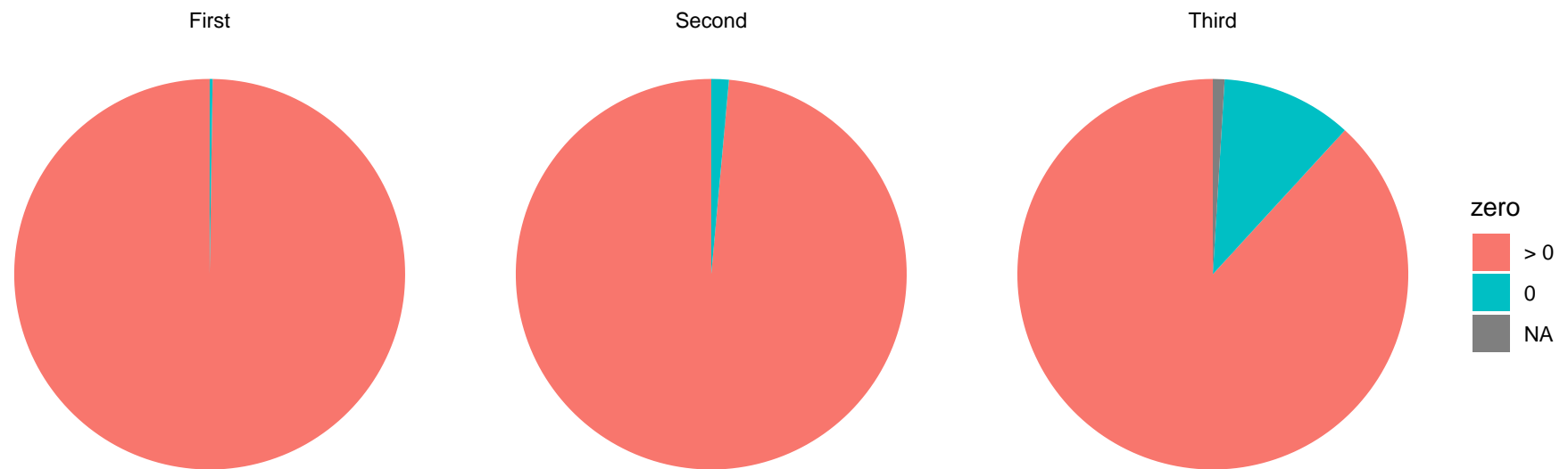
Warning: Removed 1 rows containing missing values (geom_col).

Figure 2: My caption



1.1.2 Zeroes

```
## `summarise()` has grouped output by 'Dose'. You can override using the `.groups`  
## argument.
```



The column “Third” has more zero values than “First” and “Second”; but, I think, it won’t influence models’ accuracy. Also, we can see missing values for the column “Third”; in our case, missing values mean that nobody got the third jab. I suggest replacing them with zeroes.

Replace missing values.

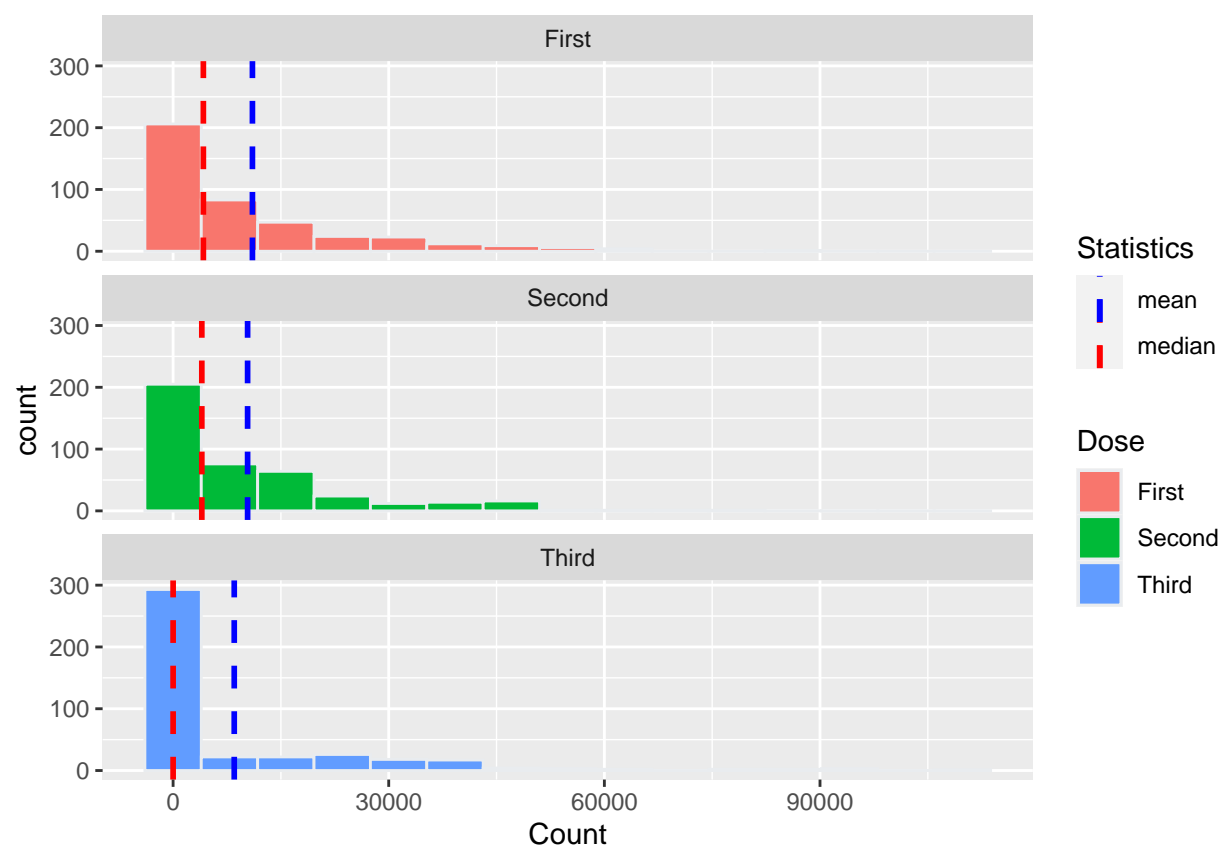
1.1.3 Data description

Median, percentiles and mean

	mean	median	Q0.25	Q0.75	Q0.9
First	11037.614	4210	2072.0	14213.0	31224.6
Second	10367.352	3998	1659.5	13708.5	28700.6
Third	8589.925	6	2.0	10629.5	33689.0

What can I say?

- Mean and median have a visible difference. So, there are large extreme values.
- For the Third dose, half of the values are below 6. That is not surprised. In the beginning, people needed to get two jabs.
- If we look at “Q0.25”, “Q0.75”, “Q0.90”, we find out that the Third dose’s wave caught up with other doses’ waves quickly. We already saw this fact on the plot 1.



Standard deviation (sd), IQR and range

	sd	range	IQR
First	13971.58	84537	12141.0
Second	13477.22	78425	12049.0
Third	17512.10	109810	10627.5

IQR and standard deviation for each dose are big, consequently, the data spread out. Also, we can see the difference between largest and smallest values in the column “range”.