ACG 7848

Group Assignment

Group 6

**Vaccine Adverse Event Reporting System (VAERS)**

**April 23, 2022**

**Xiangyue Duan**

**Qing Yu**

**Narottam Jajodia**

Business Question:

When it comes to potential adverse events, are all COVID19 vaccines ‘created equal’? Is vaccine quality constant over the different lots (batches) that each manufacturer produces?

Data Sources:

<https://vaers.hhs.gov/data/datasets.html>

* 2021VAERSDATA.csv – description of the case
* 2021VAERSSYMPTOMS.csv – symptoms (may be multiple records)
* 2021VAERSVAX.csv – description of the vaccine (disease, manufacturer, lot number)

Data Preparation:

The datasets have information on vaccines for viruses other than COVID-19. We used code to **filter out** the extra information that wasn’t useful for our purposes.

Analysis:

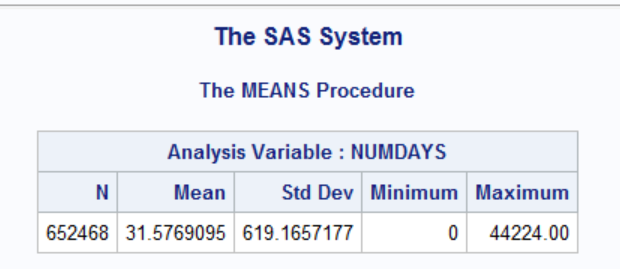
Question 1:

**Examine the number of days between vaccination and the date of the adverse event.**

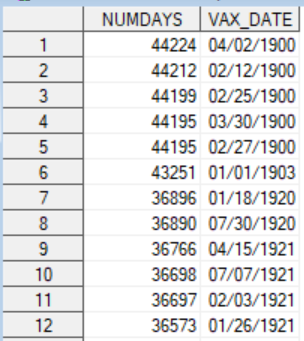
The variable ‘numdays’ in file ‘2021VAERSDATA.csv’ is a list of the number of days between vaccination and the date of the adverse event.

*Screenshot from data guide* [*https://vaers.hhs.gov/docs/VAERSDataUseGuide\_en\_September2021.pdf*](https://vaers.hhs.gov/docs/VAERSDataUseGuide_en_September2021.pdf)

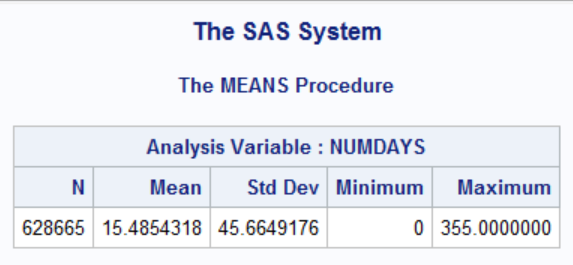
We analysed the numdays variable and found the maximum value (44224) to be very high, considering that COVID-19 vaccines were first administered only in December 2020 (NCSL, 2020).



We then looked at the vax\_date variable (date when COVID-19 vaccine was administered to a particular patient) and it surprisingly contained dates from several years ago, which was possibly driving the high numdays.

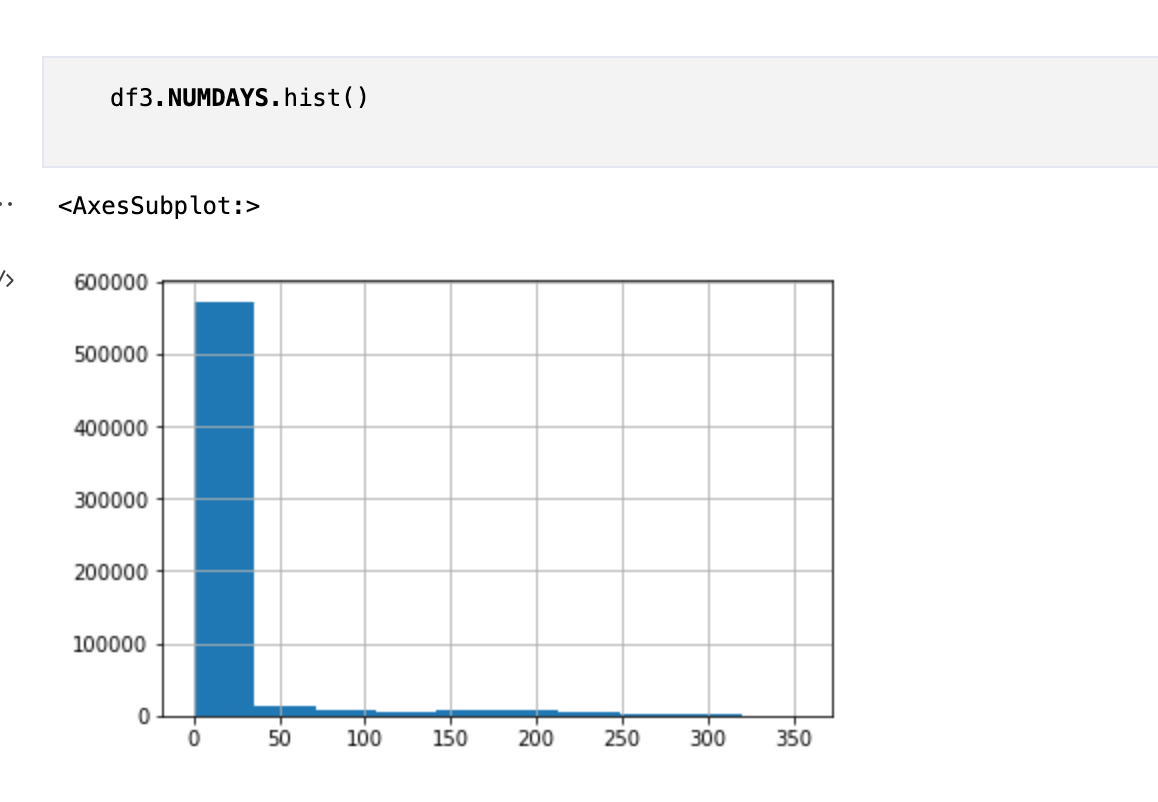


We filtered vax\_date to display only 2021 dates and then looked at numdays again. Now the values looked correct.



### Visualization:

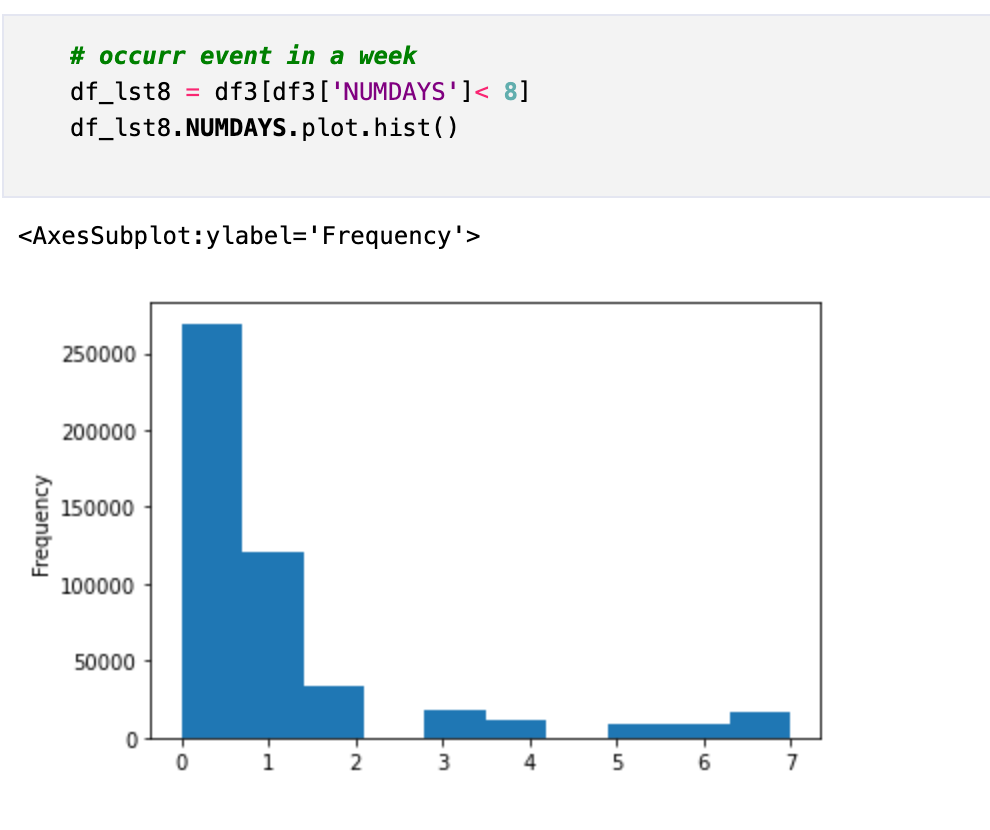
To examine the distribution of days with adverse events, we created a histogram:



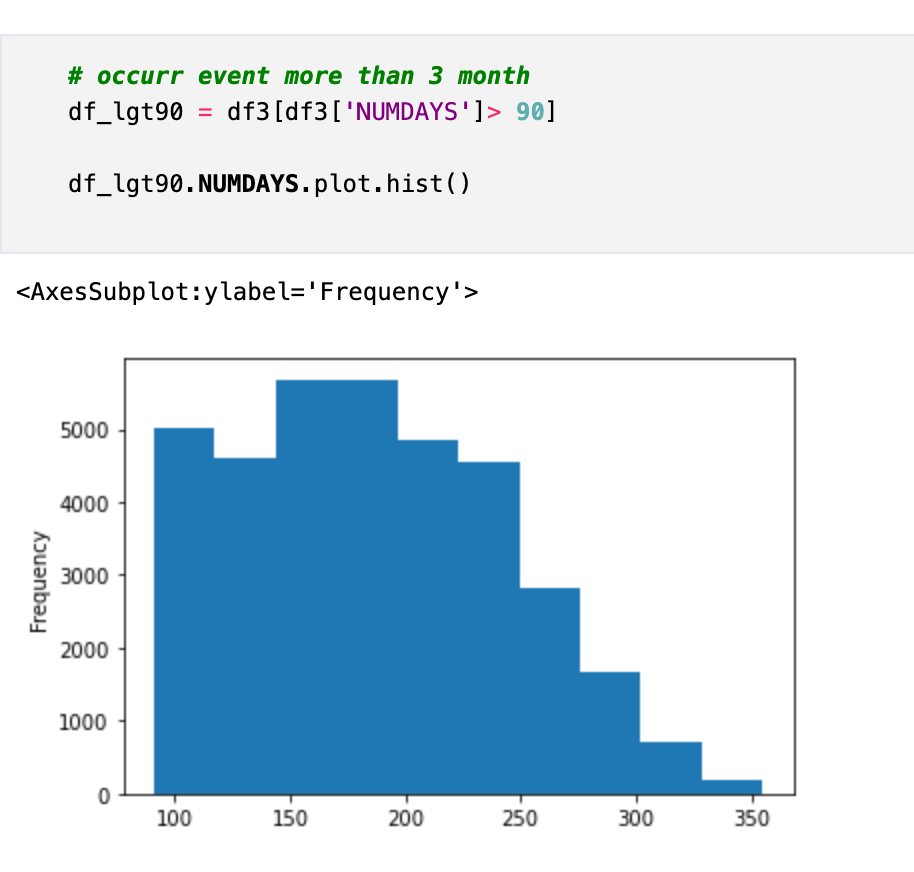
The x-axis is the number of days from vaccination to onset of symptoms. Y-axis is the number of events.

Seeing through the histogram of all *numdays* and description analysis above, both analysis shows that most of the adverse events occur during the first few weeks of vaccine administration.

To further explore, we create two histograms for the adverse event that occurs in a week and the adverse event that occurs after 3 months.



The x-axis is the number of days from vaccination to onset of symptoms. Y-axis is the number of events.



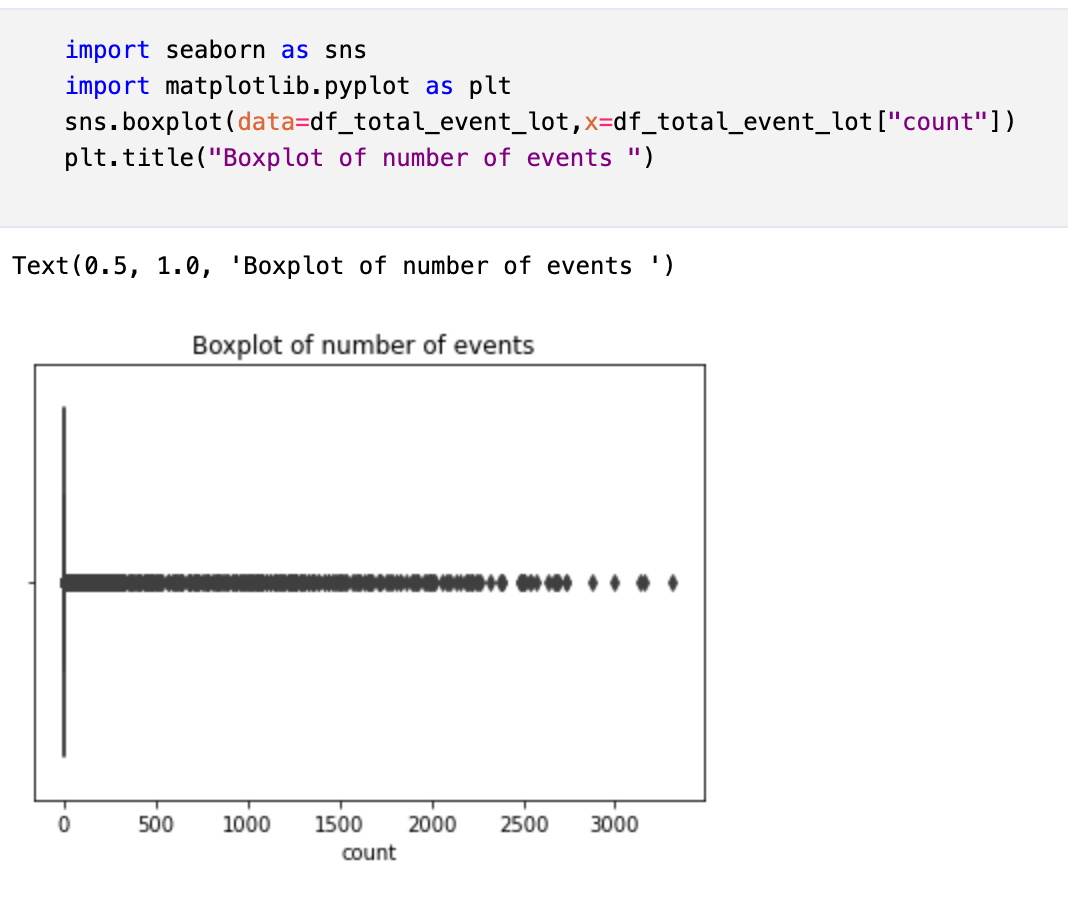
The x-axis is the number of days from vaccination to onset of symptoms. Y-axis is the number of events.

The histogram for the adverse event that occurs in a week shows a pattern that most adverse events happened immediately after vaccination.

Question 2:

**Vaccines are produced in batches/lots. Investigate if the adverse events for covid19 vaccines are clustered in certain lot numbers (as opposed to adverse events being unrelated to lot numbers).**

To investigate if the adverse events for covid19 vaccines are clustered in certain lot numbers, we created a strip plot.



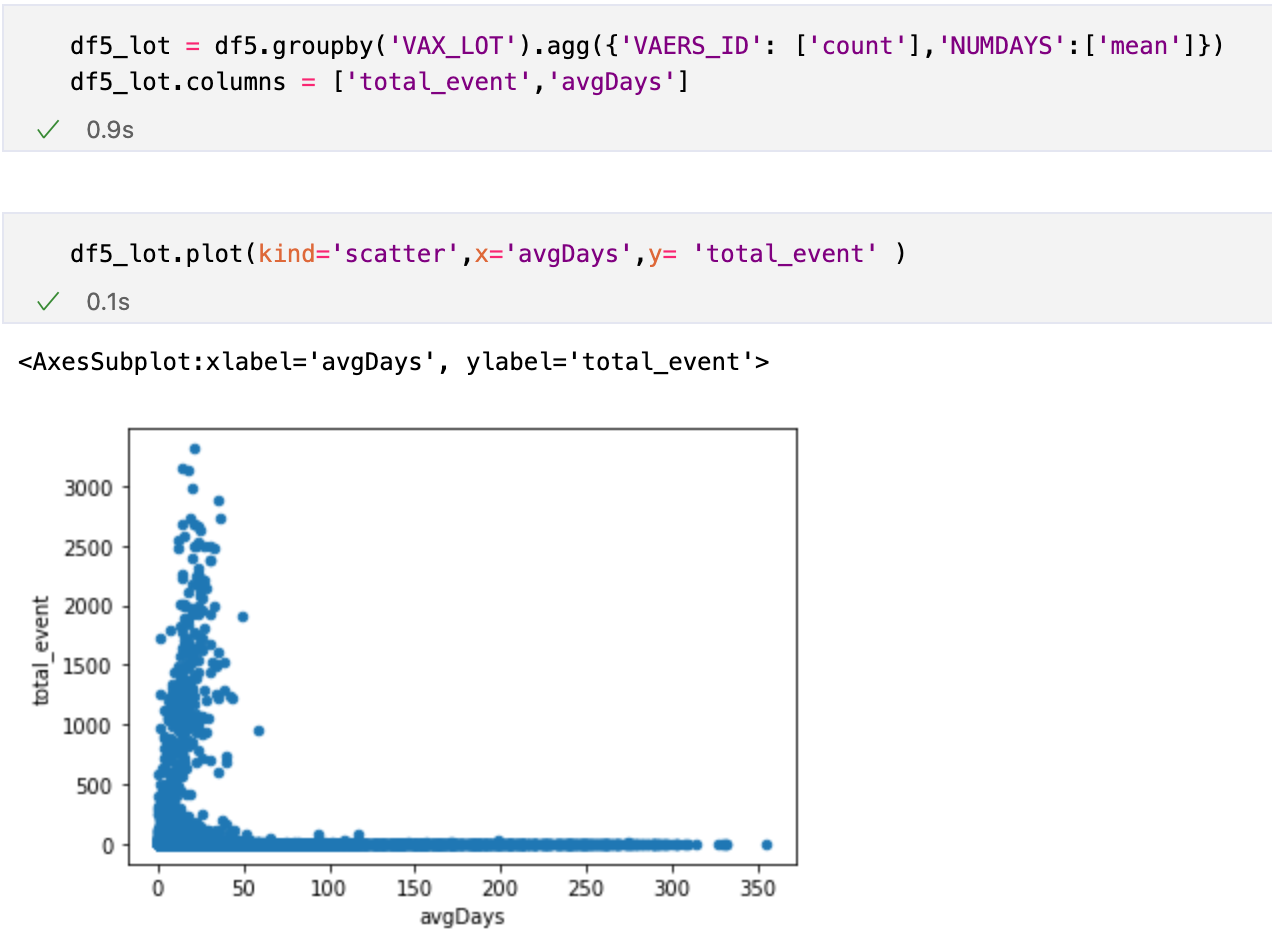
The x-axis is the number of adverse events and the circles are the individual vaccine lots.

It is clearly visible that certain lots are associated with a high volume of cases. Thus we can reject the assumption that adverse events are unrelated to lot numbers.

Question 3:

**Examine if results for part A are different for lots produced that have higher/lower adverse events (part B).**

To examine if our analysis of numdays in question 1 has a relationship with the findings in question 2, we created a scatter plot.



The x-axis is the average number of days (between vaccination and onset) per vaccine lot. The y-axis is the total adverse events per vaccine lot.

It can be observed that the majority of event groups, of all values - low to high, are concentrated between 0 and 50 average days. There is no corelational pattern. We can therefore interpret that the number of days (between vaccination and onset) is different for lots produced that have higher or lower adverse events. In other words, there is no distinct relationship between the two variables.

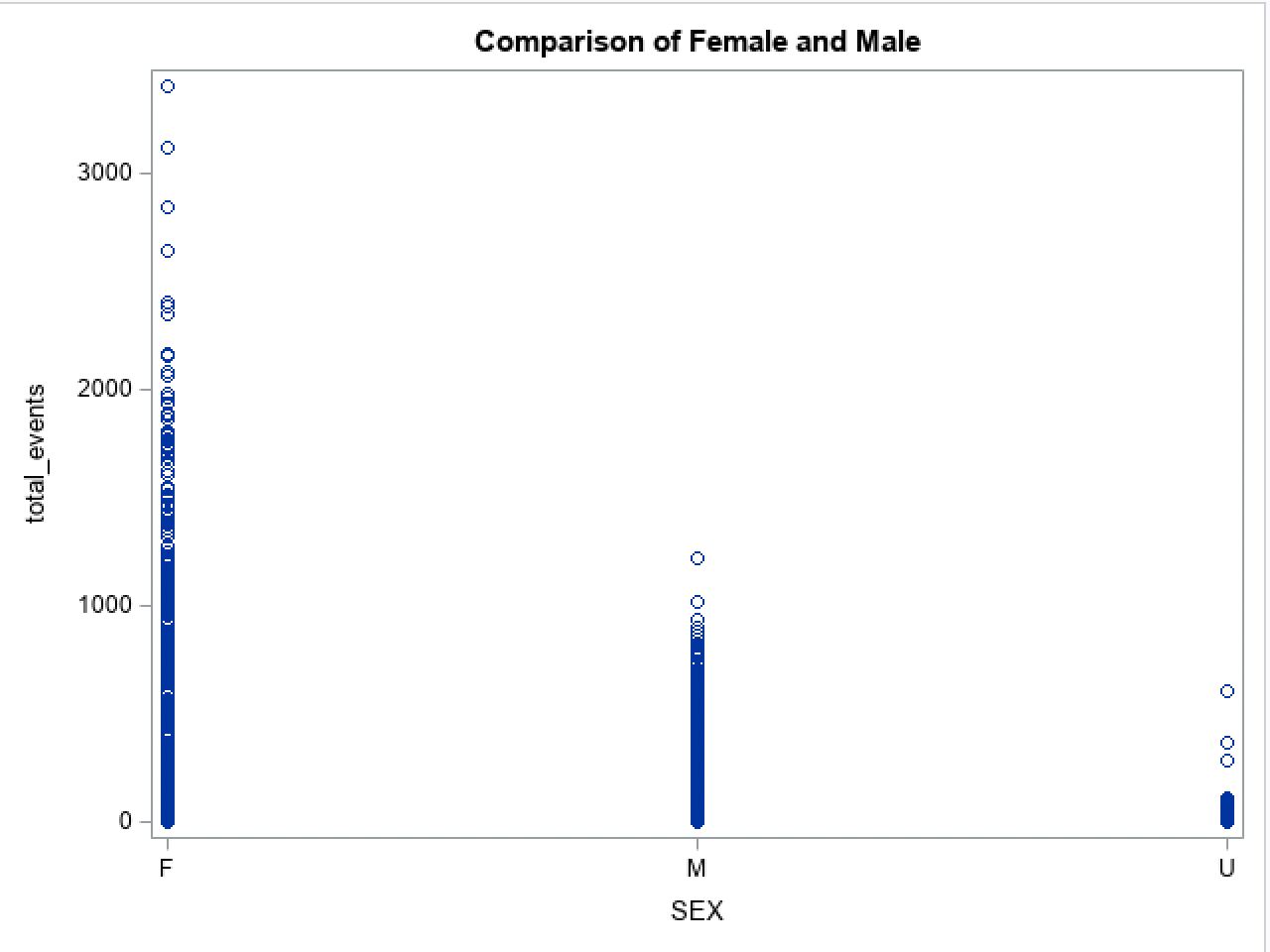
Question 4:

**Explore the data. Create several (4 or more) interesting tables/graphs.**

To further explore the data, we chose some other factors to have further analysis.

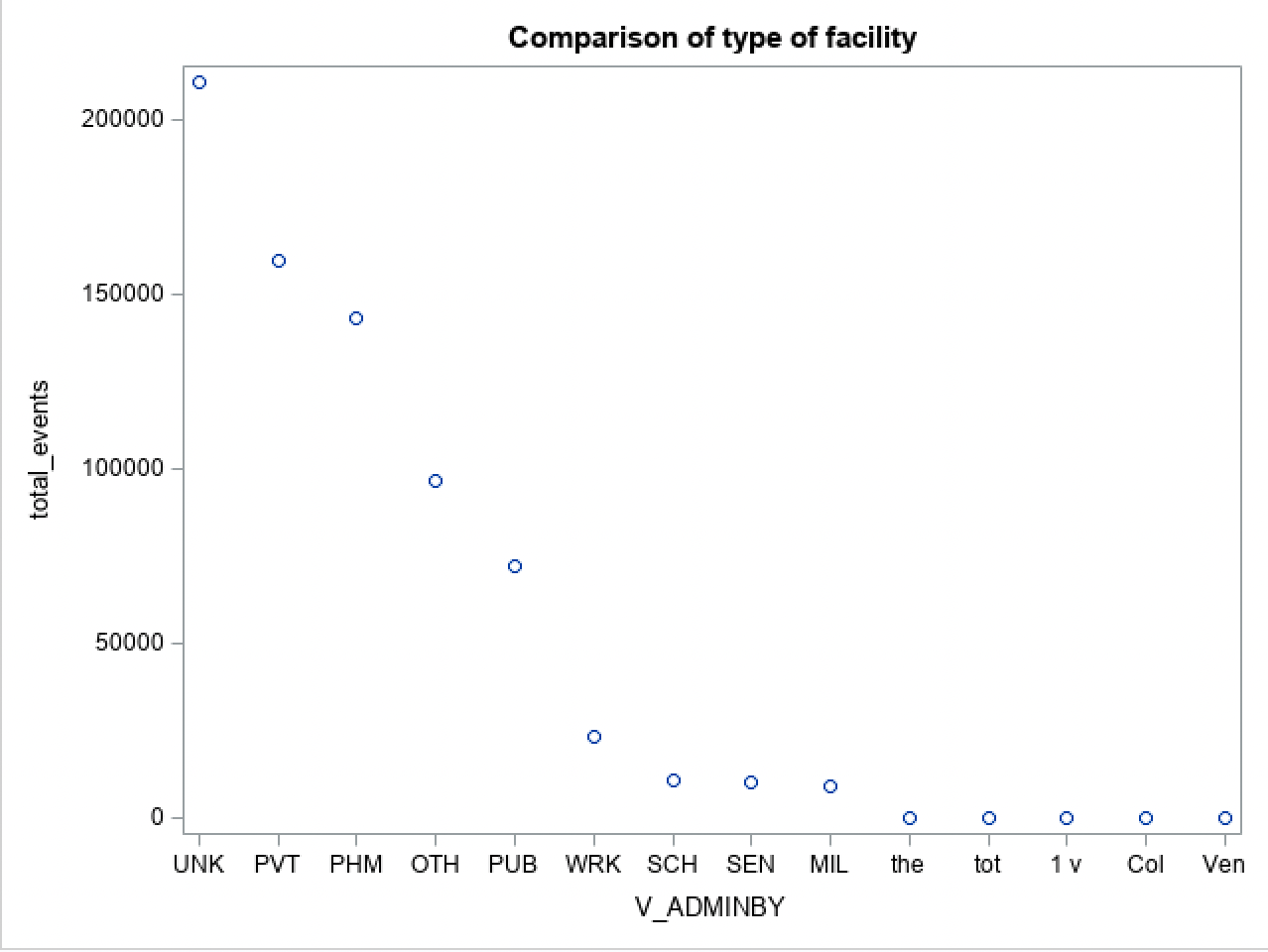
1. SEX: We chose sex first to see whether there is difference for adverse events between female and male.

From the chart below, we can see that females are more likely to experience adverse events than males. This difference may be caused by the different body structures of females and males.



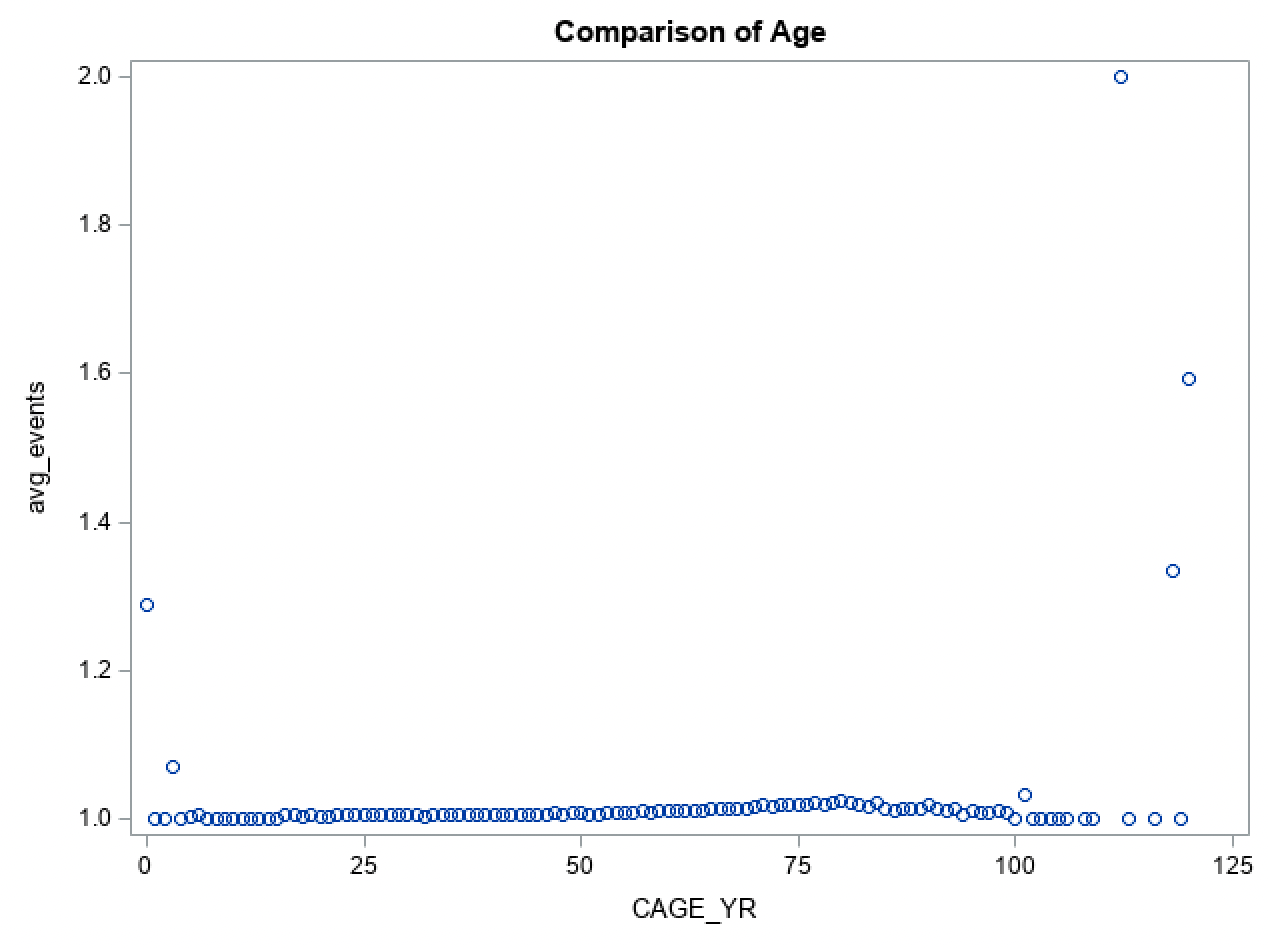
1. Type of facility: Then we focus on the type of the facilities used by them.

From the scatter plot below, we can see that it indeed has an effect on the number of adverse events. The UNK(probably unknown) has the highest adverse effect which is nearly 50000 more than the second one PVT(probably private) which may be caused by the design or other factors of facilities. This may help us to choose the best facility for vaccination.



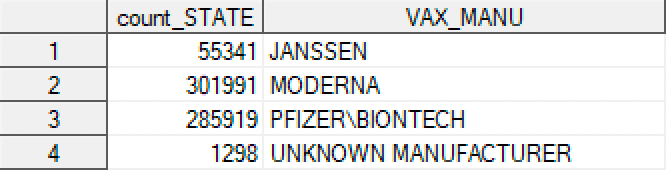
1. Age of people: Are older people more likely to suffer from the adverse effect? We made a scatter plot to explore it.

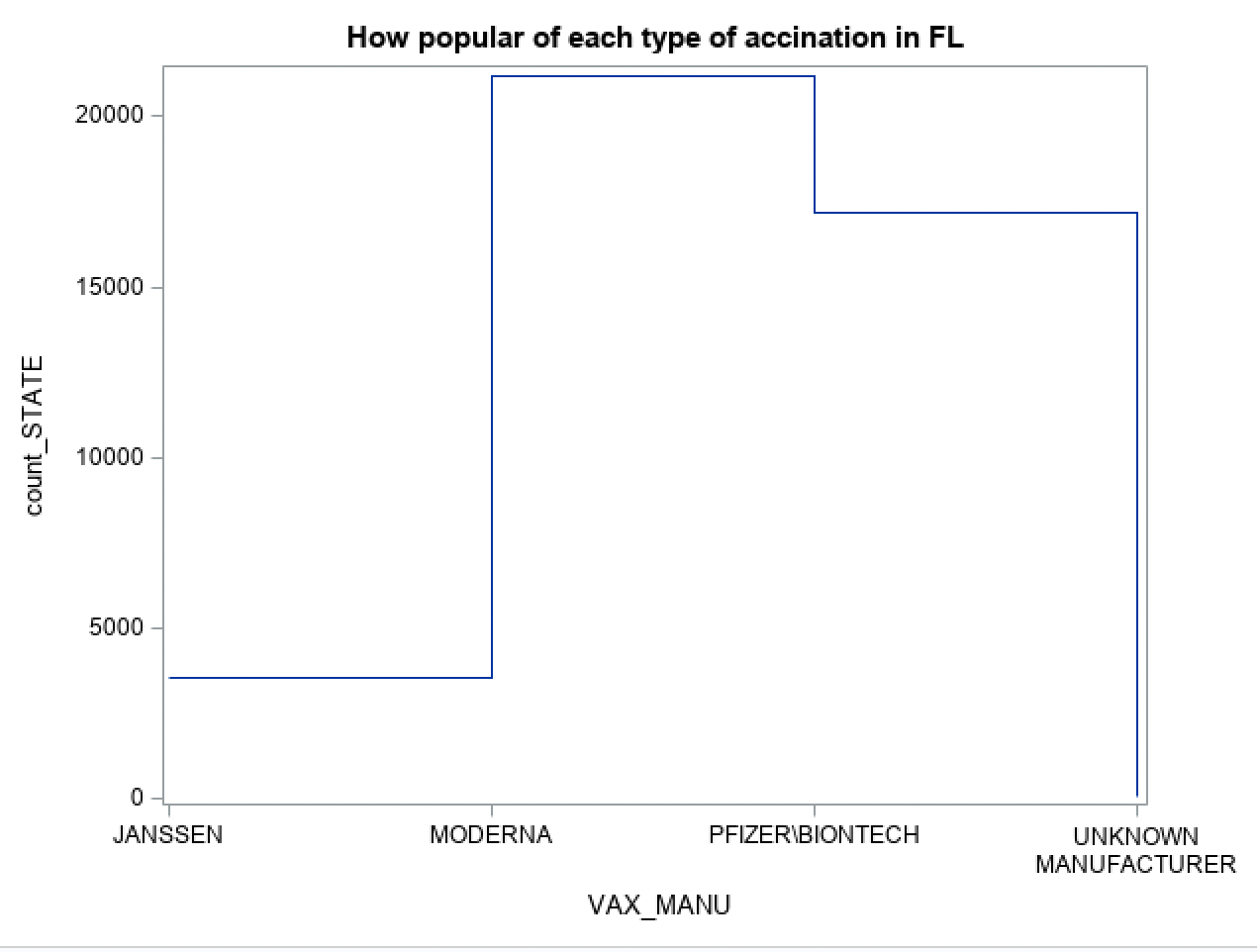
From the scatter plot, we can learn that the effect age has is not so big. As the age increases, the possibility of adverse effects will slightly increase after about 60 years old. It may be because health generally deteriorates with age.



1. Which type of vaccination is popular in all states and FL?

First we calculate the number of different vaccinations in all states and the outcome shows that MODERNA is first with 301991 and PFIZER\BIONTECH is second with 285919. Then we calculate the data in FL.



From this step plot, we find the same outcome for all states. MODERNA is the first and PFIZER\BIONTECH is the second. It may be because Americans think the MODERNA and PFIZER\BIONTECH are more reliable and safe or that the size of these two manufactures are bigger than other COVID-19 vaccine manufactures in America.  


**END**