1. Give an example of Missing at Random and not missing at Random data. Explain your answers.

MAR: Defined as having a situation where the missingness of data can be explained by other observed variables in the dataset, but not by the missing data itself. The missingness of attending tutoring (the variable with missing data) might be related to the students' test scores and their gender. For instance, students with lower test scores might be more likely to attend tutoring, and this could vary by gender.

NMAR: Defined as having a situation in which the missingness of value depends on the unobserved value of itself even after looking at the observed values. Eg: The data of accidents.csv provided to us which has a column 'Age of Casualty' which is a NMAR because the missing values cannot be determined by looking at the other values present like (Class, severity and others).

2. Using functions from the dplyr to delete any unnecessarily columns or rows. Justify your selection in the report. Reasons could be duplication, non-changing value.

The unnecessary column like Local.Authority was removed as it had only one value and other column Daylight.Dark was removed as it had the same feature as the Lighting.Condition column.

```
Local.Authority
              1
```

Lighting.Conditions <chr>	Daylight.Dark <fctr>
Darkness: street lights present and lit	Dark
Daylight: street lights present	Daylight
Darkness: street lights present and lit	Dark
Daylight: street lights present	Daylight
Daylight: street lights present	Daylight
Darkness: street lights present and lit	Dark

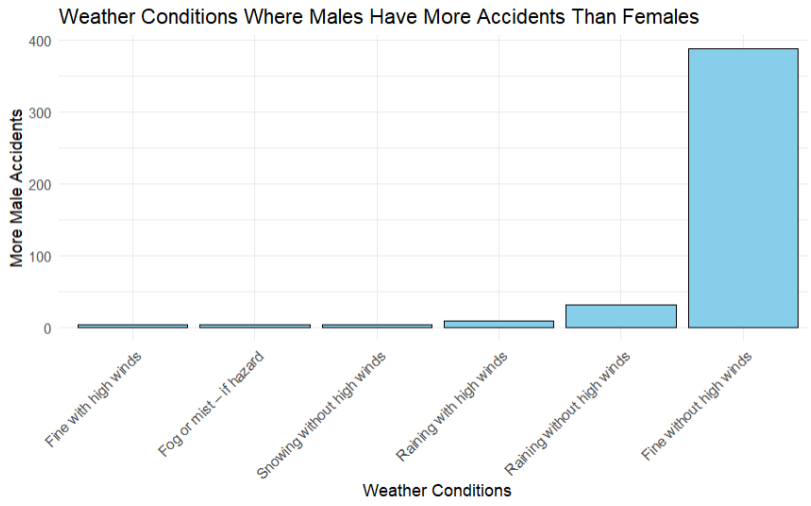
3. Examine "Age of Casualty" for any outliers using the three methods discussed in the lectures. Compare the output of these methods. Which one is the best? Justify your answer

Three sigma rule is the best among the three.

Exploration

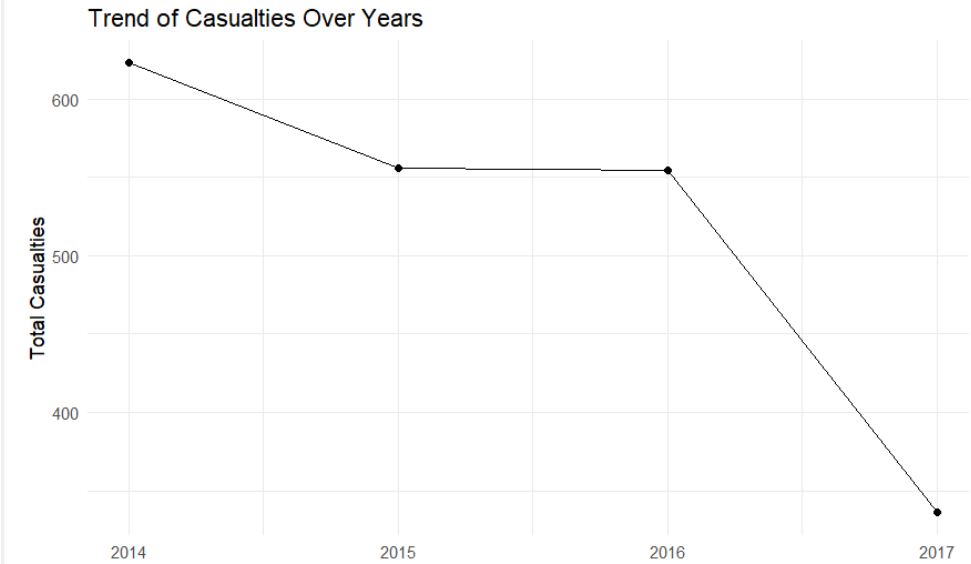
4. Is there any weather condition where male drivers/riders have more accidents thanfemale drivers? Print out how many (e.g., "male drivers have ----- more than females when weather is -- --.

   Yes, there are 6 weather conditions where male drivers/riders have more accidents than female driver. They are: Fine with high winds, Fog or mist-if hazard, Snowing without high winds, Raining with high winds, Raining without high winds and Fine without high winds (highest).



5. Is the number of casualties increased or decreased over time? Which year has the highest number of casualties?

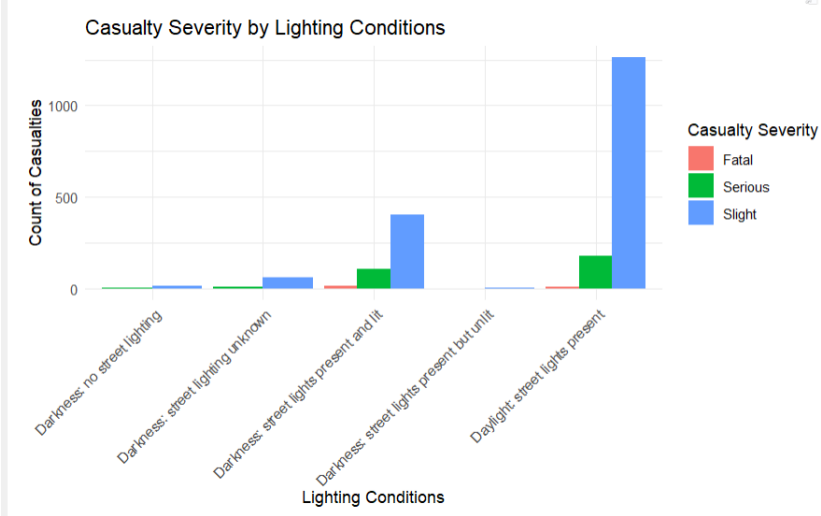
   The number of casualties have toned down a lot compared to that at the beginning. The year 2014 had the highest number of casualties.



   The reason for the downfall of number of accidents may be due to the implementation of new and stricter traffic rules as well as the increase in safety features in car.

6. Draw a plot to explain the relationship between the following:
   • Light conditions and severity
   •Weather condition and number of vehicles involved

• Light condition and severity



Darkness with no streetlight:

There is no fatal injury and only few seriously injured with a slightly higher number minorly injured

Darkness: street lighting unknown

Although no fatal injuries, the number of seriously and slightly injured has tripled compared to darkness with no streetlight.

Darkness: street lights present and lit

The lit streetlight at darkness has the highest number of fatal injuries (15), 107 seriously injured and 405 slightly injured. The reason might be because of low volume of traffic at night and well lit streets.

Darkness: street lights present but unlit

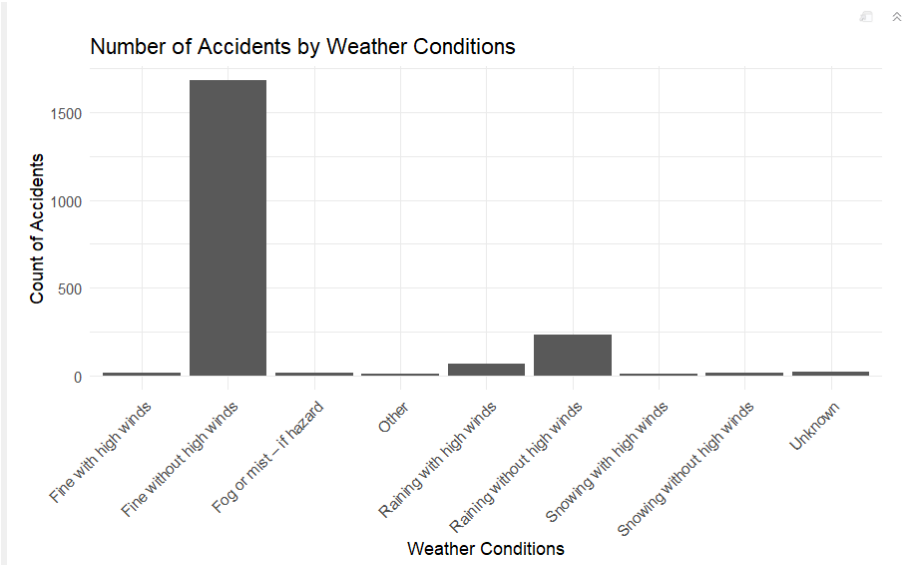
The unlit streetlight in darkness has the lowest number of injuries overall(0 fatal, 1 serious, 3 slight). The reason might be because of although low traffic, but low visibility as well.

Daylight: street lights present

Shows the highest number of accidents in every category except fatality(10 fatal, 178 serious, 1263). The plausible reason might be high volume of traffic and better visibility of accidents happening on the broad daylight.

Other 2 conditions (Daylight: no street lights present and Daylight: streetlight unknown) has 0 number of casualties as the obvious reason being no streetlight being lit during the daytime.

• Weather condition and number of vehicles involved



Fine without high winds:

The weather condition has the highest number of casualties reaching almost 2000 and is exponentially higher compared to the next in line (Raining without high winds). The likely reason might be high volume of traffic

Raining without high winds:

Although not as severe as fine without high winds, the number of vehicles involved in accident is noticeable (250). The most probable reason cold be more cautious driving during rain as the tires may slip.

Raining with high winds:

The third condition with the highest number of vehicles involved in accident but incomparable to aforementioned ones (100). The probable reason could be even more cautios driving than the rain without winds as there are 2 factors to look at.

Others (everything except aforementioned):

All the other weather conditions have less than 50 vehicles involved in an accident.