AmirHossein Mohammadi

810197689

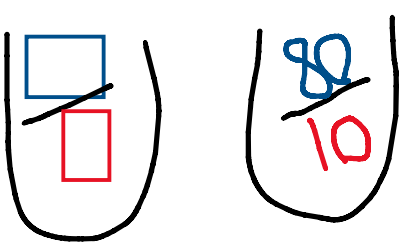
Question 1:

a. During summer, eating ice cream is more pleasant than other times of the year. On the other hand, temperature has also positive correlation with murder rate since in summer days people go outdoors more often. So, temperature is confounding variable.
b. Food is confounding variable. Food consumption increases as child birth increases, and food wastage also increases if more food being reachable by humans. Storks are always looking for land with more food so their number increases.

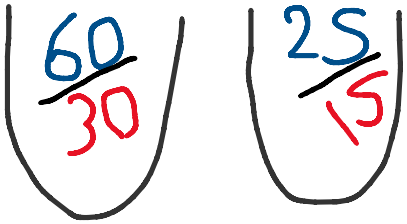
Question 2:

Italy has more CFR in each group than China, but it can't be concluded that Italy has more CFR in aggregate. For further illustration, conditioning probability in groups gives us perspective that could not necessarily be true for aggregate probability. For instance consider example below:

We have two pairs of candy jars, and each jar has two types of candies blue candy and red candy. We like the blue candy more than the red one, the question is that if we have more percentage of blue candies in left jars, will be also more probability of picking blue candies from aggregate of left jars or not?

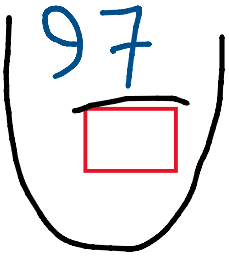


$\frac{37}{40} > \frac{80}{90}$

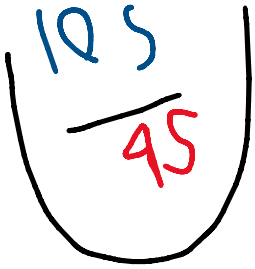


$\frac{60}{90} > \frac{25}{40}$

In both pairs left jars has probability that we pick blue candy, but in total left jars doesn't have more probability of picking blue candy.



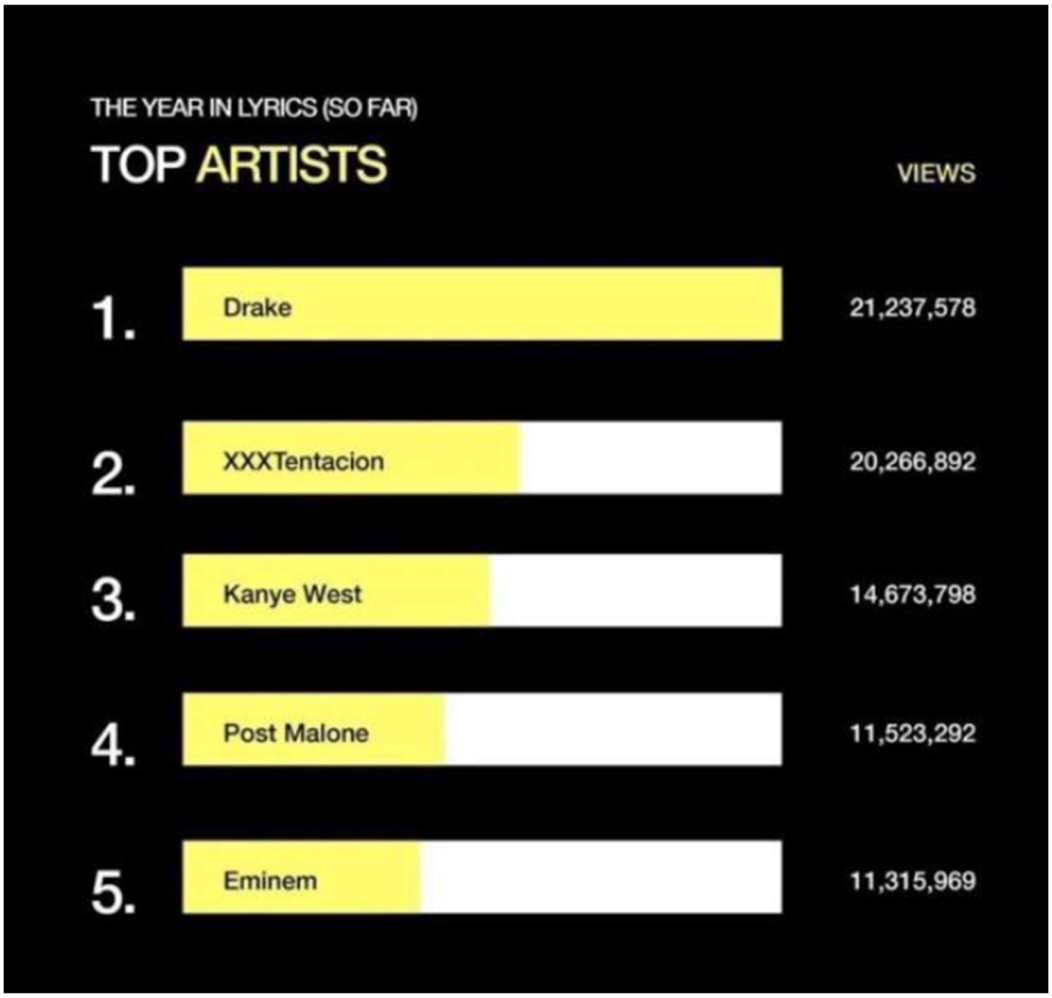
$\frac{97}{150} < \frac{105}{150}$



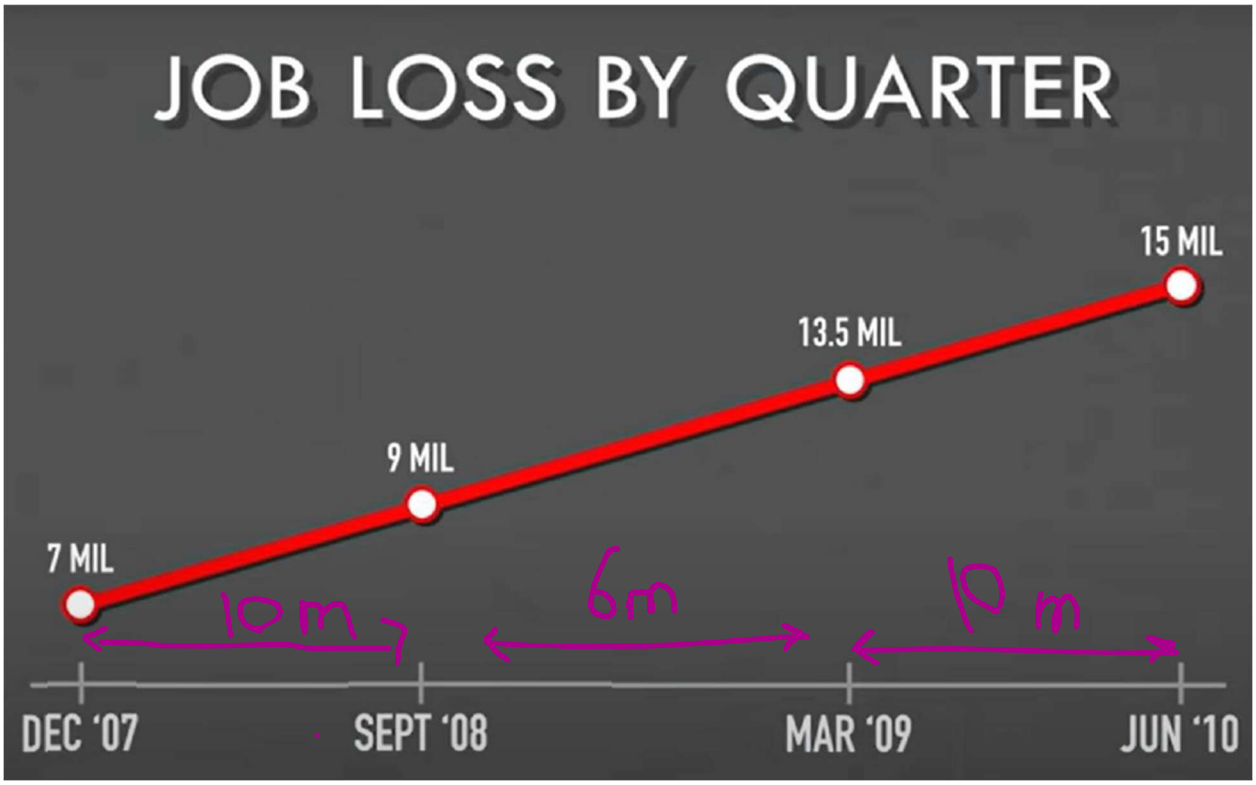
This example can be extended for more multi group. Therefore, from above we can conclude that the Ali's statement is False.

Question 3

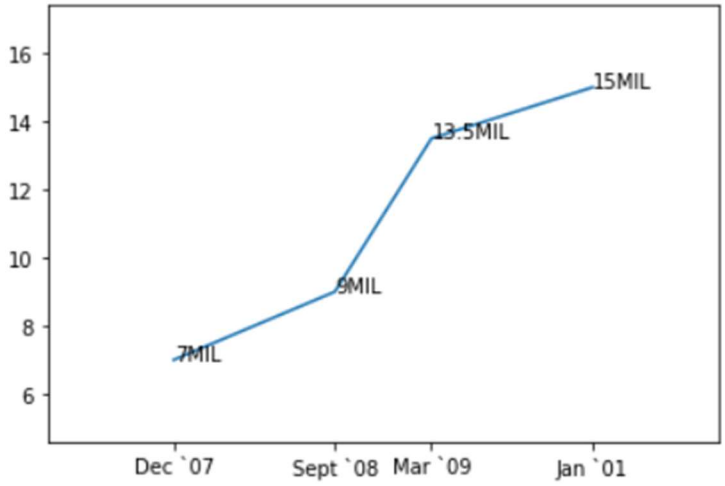
a. As we can see Drake has 21M viewers and XXXTentaction has 20M viewers and their difference is 1M, but in the bars XXXTenction is shown about half of Drake which is not true according the views. And this has been repeated for other rows too, we can see Kanye West has 6M less views than XXXTenction but there is a little difference in their bar plot.



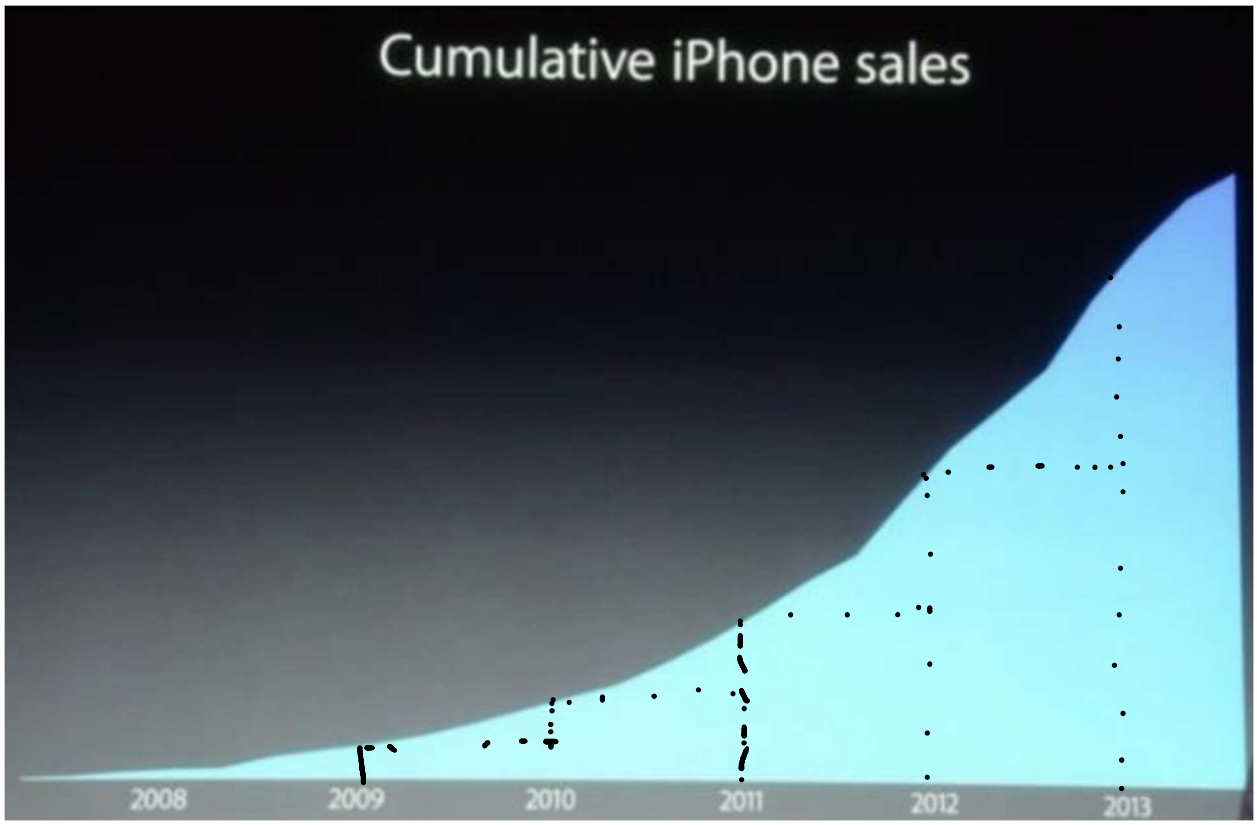
b. The plot tries to hide the fact that so many people has lost their job from September 2008 till March 2009. It tried to hide this fact by expanding the x-axis to try the induce that idea “Although 4.5 million people had lost their job, it also took more time”



The real plot is like should like this:



c.



In this Picture Tim Cook tried to induce the idea that iPhone sales has been increased exponentially, but the plot is Cumulative iPhone sales. We can see from drawn dash lines that the iPhone sales does not have exponential trend.

d. This graph tried to induce the idea that Annual Global Ocean Average Temperature has not changed a lot since 1880. But if we look closer, we can understand that Global Ocean's water volume is a massive amount of water and increasing a little bit in them needs a huge amount of energy. On the other hand, livings are very dependent to small changes in water temperature and only small change in water temperature can ruin their life. So, it would better if y axis was limited into 15 - 18 degrees and its redundant parts was omitted.

Question 4:

Two independent variables has influence on generalizability and causality. The first one is random sampling from population. Since our sample should be representer of our population we should have a good amount of random picked sample from distribution. The amount of needed data is dependent on the task.

The other important variable is the experiment setup. It should be setup in a way that no other confounder influence our response variables. Appropriate experiment can lead to a correct causal claim. So the table is as the following:

		Generalizability	
		no	yes
Causality	no	Experiment setup	Random Sampling
	yes	experiment setup	Random Sampling

Question 5:

a. Since IMDB aim is to rank the movies despite their nationality, and due the fact that The Heroes is a favorable TV show for many Iranian children, it is obvious that Heroes score on IMDB has bias because of Iranian culture.
b. As old saying goes ‘dead men say no tales’ . The one who died in conflict can not defend himself, on the other hand the other party can gather evidence in order to get rid of accusations.
c. Taxi driver is not making a causal claim. Traffic and police presence have correlation but they do not have causation.
d. Since classes with higher population will be more probable to sampled it has biased toward classes with more population.