

Opinion reversal using information visualisation

Sumit Chawan 18200549

Proposed Hypothesis

The proposed hypothesis mainly focuses on the question **“Can biased visualisation techniques alter the audience's choice of a political party and can we present correct data in misleading form so that it makes audience reverse their choice?”** Various visualisation techniques are being used by the experts to provide data insights which play a significant role in our day to day lives. Since the possibilities for visualising data are endless, we have decided to focus on the political domain. In our experiment, we are trying to show how manipulating the axes can mislead the audience into believing something which is completely false. With social media being in everyone's pocket now, the risk is higher than ever, as people see everything in minimal time and are more prone to misinterpreting the information. In other words, seeing has become believing.

Our topic addresses politics in general, as it deals with the development indices of the country, such as unemployment rate, GDP and FDI. A small tweaking in these highly sensitive indices using deceptive representations may result in a catastrophic change. These changes, being either intentional or non-intentional may lead to misinterpretation of data, are termed as visualisation lies.

The prime importance of conducting this experiment is to understand the threshold up to which the skewness created in data leads to Opinion Reversal amongst the audience. This technique of controlling and regulation of information being provided to the people is actually known as ‘Impression Management’. This topic has a very wide scope, so we have narrowed it down to various axis-manipulation techniques like Inversion of Axes, Truncated axis etc. that actually make use of Impression Management and mislead the audience.

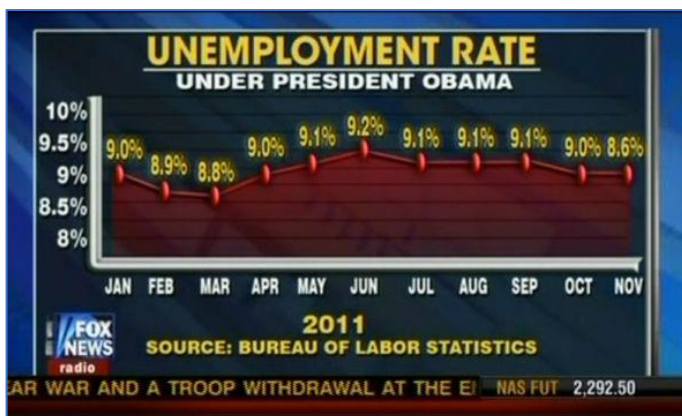


Figure 1.

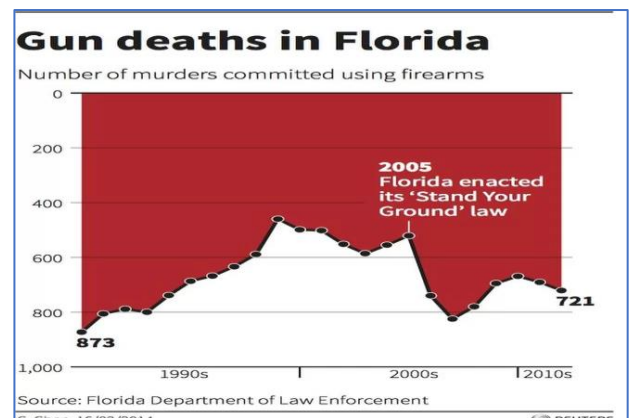


Figure 2.

For instance, in 2011 there was a news on FOX NEWS channel regarding unemployment rate in USA. If we glance at the graph in Figure 1., it looks like after a small drop in unemployment, it starts to go up and gets worse under Obama's presidency. But if we closely observe the graph, we can easily see that the graph has been manipulated in following two ways:

- Firstly, the Y-Axis is truncated by initializing it from 8% rather than 0%. This makes it difficult to make a distinction between exact figures which are actually close but are exaggerated in the graph above.
- Secondly, if we observe the unemployment rate in November, we can see that its 8.6% but, on the Axis, it has been depicted almost near to 9% which creates ambiguity and gives a misleading information(Phenomenon of compressed axis).

In the above scenario, the unemployment rate is actually decreasing but the opposite has been depicted by inappropriate axis selection.

The Area chart in figure 2 represents the gun deaths in Florida during the Governance of Jeb Bush who was the Republican candidate. The article from Reuters, A news Agency in USA inverted the graph which leads to a conclusion that the gun deaths have increased during the term of Jeb Bush, but on viewing it carefully, we can perceive that gun deaths have actually gone down in the state from 873 to 721. But this is not evident due to inversion of axes. This example is exactly in alignment as to how manipulation techniques can have a reversal of opinion from the perspective of general public towards the current governing body.

In order to complete the experiment, we will be making use of web form to conduct a survey on target audience who would be centrists who are fact and number oriented (conscious mob) rather than emotional ones.

1. Experimental Method

2.1 Overview

The suggested plan is to identify different axis manipulation techniques to confuse people into believing the misleading data. For each scenario, we will be randomly showing two graphs, one of which will be manipulated and the other will be correct. We have 9 scenarios, which means that a total of 18 questions will be displayed to the participants.

There are plenty of techniques available to manipulate the audience views, out of which, the below methods for axis manipulation will be used in the scope of the experiment.

- Inverted Axes
- Compressed Axes
- Cherry Picking
- Truncated Axes
- Cumulative Axes

While conducting the experiment, the dependent variables will be speed, accuracy and precision because they are correlated with the type of responses given by the participants and the independent variables will be the different graphical visualisations which are made using the above axis manipulation techniques and also that with correct representations

Confounding Variables are the variables that are often responsible for damaging the internal validity of the experiment due to the researcher's inability to control or eliminate them. The results may show a false correlation between the dependent and independent variables, leading to an incorrect rejection of the null hypothesis. Since all the participants

are human who may have different perspective towards data, we could see the below possible confounding variables:

- Age
- Colour blindness

The effect of these confounding variables is nullified by selecting the audience which fulfil the below listed criteria.

- None of the selected candidate is colour blind
- The selected candidates fall between the age group of 20 – 60 years as the visibility tends to decrease with the increment in age.
- Participants are expected to have a basic knowledge of numbers, axes and graph interpretability.

Experimental conditions will be quite straightforward. All a participant would require is:

- A desktop/Laptop/Mobile
- A working internet connection
- An email id, to which the link of survey can be send
- Ability to draw inference from a graph.

The experiment is based on 'Within-in' a group approach, which will facilitate to interpret the reversal of opinion within a group of participants. The experiment also aims at providing the participants with enough visualisation using various graphical representations that will help to analyse the results for various scenarios and support the hypothesis.

2.2 Data Collection

The survey consists of different axis – manipulation techniques, having 9 pairs of questions and all the responses to these 18 questions for each participant are stored on the server which will be exported as excel spreadsheets for performing further analysis.

While conducting the experiment, a mix of subjective and objective data is collected from the audience.

- Speed – The time taken by the participant to answer each question and all the questions.
- Accuracy – Recording all the correct and wrong instances of the answers for calculating the percentage of opinion reversal.
- Interpretation ability – To assess how well the participant understand the data as our aim is to manipulate the graph.

The reason for collecting the above facts is that our assumption is that the lesser is time taken to answer, higher is the probability of selecting the answer based on the first impression about the graph.

Objective measure taken is the assessment of time taken by the participant while answering the question and the subjective measure is the interpretation (answers) made about the graph by the participating candidates.

The data collected from the questions will give us an idea if people are actually getting confused by skewed data or not. With the collection of all the responses we will create a table which will help us in retrieving the percentage of opinion Reversal.

2.3 Selected Subjects

For conducting the experiment, we have decided to choose UCD students of different age groups and gender. It is expected that the selected students have at least a basic understanding of what a graph looks like and how to draw a conclusion from a certain graph. Having additional knowledge is good but not necessary.

The students were selected because it is convenient to conduct the experiment if the audience already have little understanding about what is expected from them. Also, being a college student, each of them will have an e-mail account which is must in our case as the link for the completing the survey will be sent on their e-mail addresses. The last reason was that college students need minimal or no support while filling a generic form.

To source the participants, two approaches were taken. The first one was conducting the experiment in closed room on our fellow classmates within stipulated time so that the responses are more prone to being incorrect. The other one was to talk to other students and ask them to volunteer for filling the form. Since the form was a short one, enough people volunteered and gave their e-mail addresses and completed the form as soon as they received the link of the form.

2.4 Data Analysis

In order to analyse the data, the first step is to collect the responses of all the participants stored on the server. This data stored on server is later exported into excel spreadsheet and can be used to perform further analysis. With the help of this analysis, we will be able to prove that people's perception about a specific graph may change if he/she is shown a manipulated or skewed representation.

For each participant, there will be a total of 9 pairs of responses (one pair being correct response for correct graph and an expected response from incorrect graph for same dataset).

Manipulated Representation	Correct Representation	Opinion Reversal
0	0	0
0	1	0
1	0	0
1	1	1

Where,

1 = Expected Response

0 = Any Random Response

Figure 3. Intersection Operation.

A pair of responses is considered to be Opinion Reversed only if we get the resultant of intersection of expected response for manipulated graph and Correct Response for correct graph as 1.

Step -1 :

The top-80 candidates who completed the experiment in least time out of total 100 will be considered.

Step -2 :

For each candidate, assign a score of 1 for every expected response and a score of 0 for any other response recorded.

Step -3 :

For each pair (correct-manipulated for same dataset), if both the scores are 1, assign a score of 1, which signifies that the opinion is reversed for a particular pair of questions. Otherwise, assign 0.

Step -4 :

Calculate the average opinion Reversal percentage by taking to consideration all the 80 selected candidates.

The main aim of data analysis is to calculate the Average percentage of Opinion reversal for the selected candidates. The higher percentage of opinion Reversal for selected candidates greater is the probability of our hypothesis being correct.

2.5 Practical Setup

The volunteers for the experiment will have to give their e-mail id after which they will receive the link for completing the survey form. An active internet connection is required for receiving the link and submission of the response. The participant is expected to give his basic details and answer all the 18 questions, each of will have 4 options. The participant has to choose the best possible option amongst all.

The experiment will be conducted online with the participant to answer all the questions as quickly as possible. To avoid any distractions, participant's will be asked to switched off their mobile phones.

The participant will be given below instructions:

- All questions must be answered
- It is a time bound test
- In case not sure, select the best possible option

The experiment will be conducted in a closed room to avoid any distractions and no printed paper will be provided. All the graphs and the associated questions will be in the survey form. Since no calculations are required, no other devices will be provided.

The survey form will comprise of 18 questions of which the 'section 1' – comprises of 9 questions generated using misleading visualization techniques as mentioned earlier and the section -2 will comprise of 9- correct visualisations i.e. the ones which represent the same data but in its true sense. All the questions are strictly multiple choice wherein the participant is allowed to choose only one correct (best possible answer) as per his interpretation of the graph. Both the sections will follow an orderly pattern of answering, the section 1 followed by section 2. This division of sections is purposely made so that the audience is not able to make out easily any relation between the manipulated visual and its corresponding correct visual. Also, the time constraint which is used creates a furthermore ambiguity to derive any relation between the each correct-manipulated pair of question.

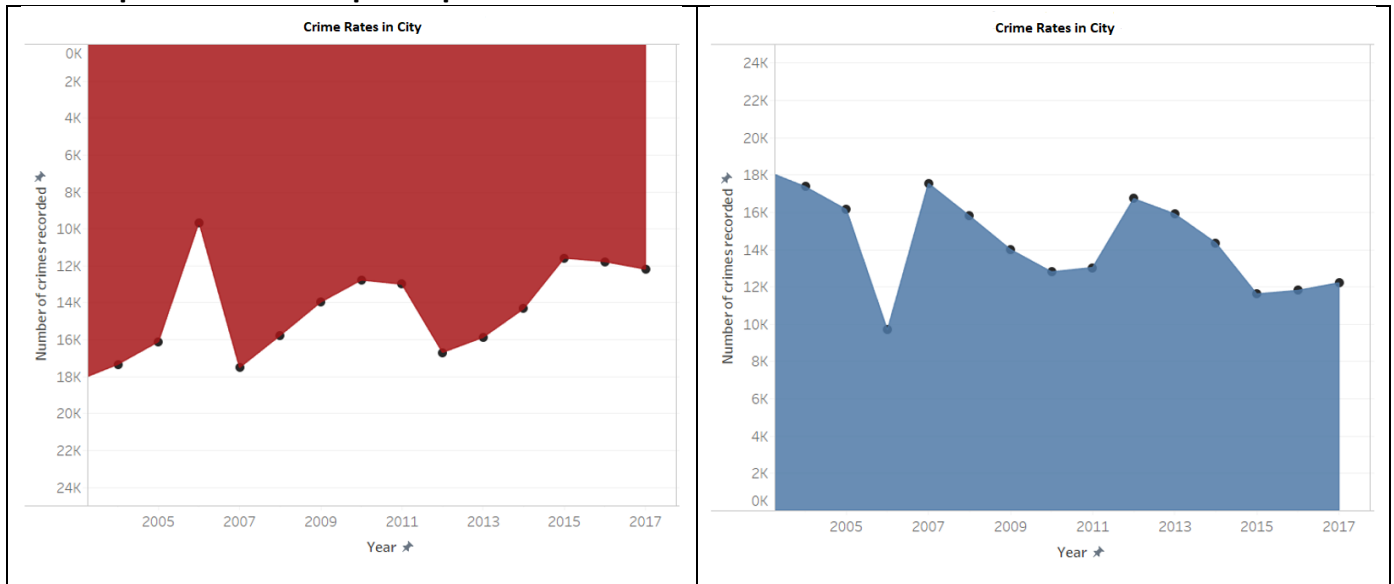
The survey is created using Google Survey forms, which facilitates the survey result storage on the server itself. Also, the survey outcomes will further be downloaded In Excel spreadsheet. The result for each participant contains the time taken by the individual to complete the entire experiment and this will be used to sort the results for top 80

participants. Only selecting the top 80 applicants based on least time is used as one of the requirements because lesser is the time taken by the participant, higher is the probability that the individual will mark the option based on the very first perception made by him on the graph.

2. Data Visualisations

1. Crime rates in city from 2005 to 2017

Manipulation Technique Implemented: Inverted Axes

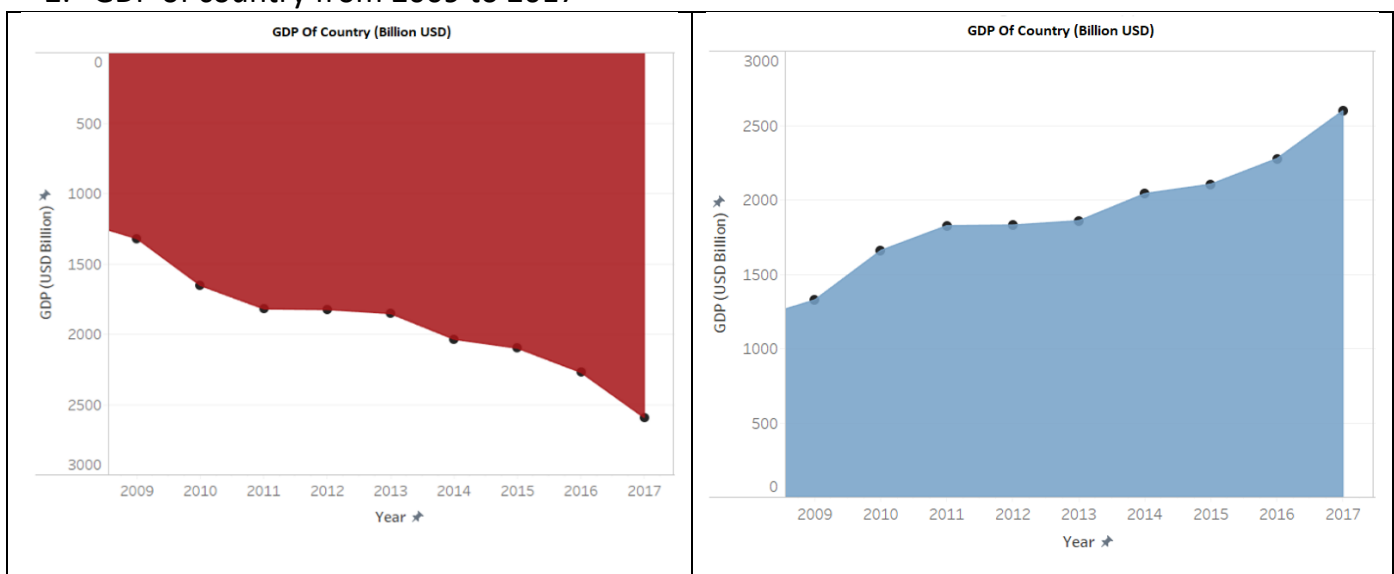


Manipulated Graph

Correct Graph

Justification: The correct graph clearly shows that there is a decrease in crime but due to inverted axes, the participants may get confused in believing that the crime rates are actually increasing between 2005 and 2017.

2. GDP of country from 2009 to 2017



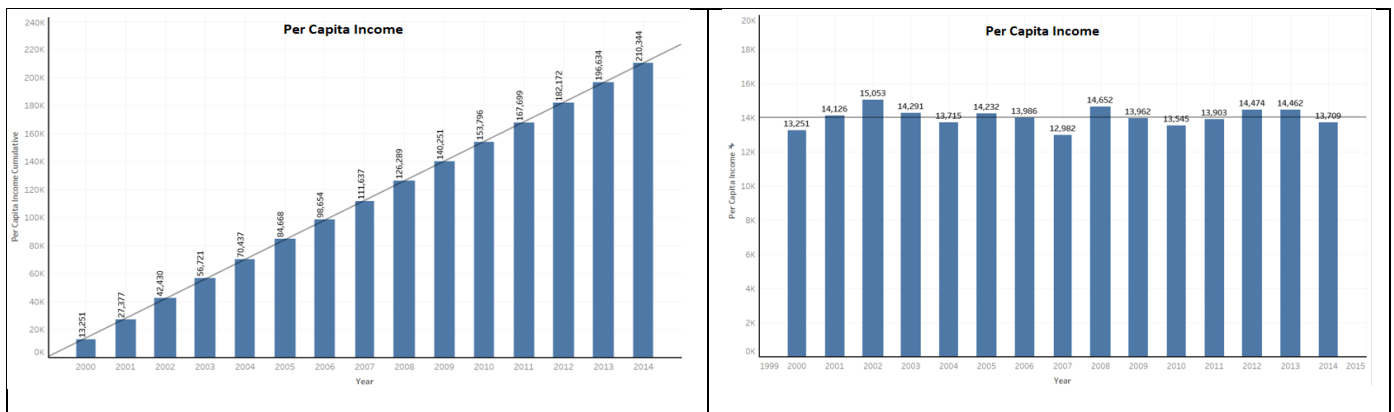
Manipulated Graph

Correct Graph

Justification: At the first sight of the manipulated graph, the audience can easily fall for believing that there has been a continuous fall in the GDP of the country from 2009 to 2017 but that is due to the axes being inverted. In actual, the GDP is consistently rising which is tremendous for the country's economy.

3. Per Capita Income for a country from 2000 to 2014

Manipulation Technique Implemented: Cumulative Axes



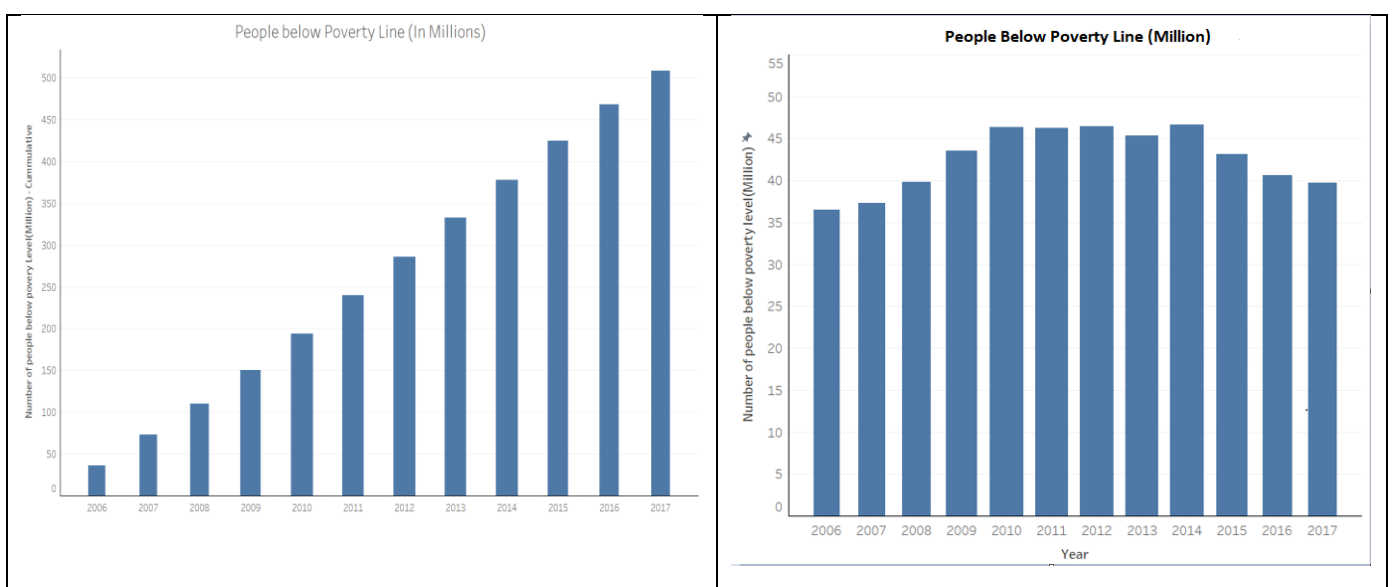
Manipulated Graph

Correct Graph

Justification: If a user takes a look at this graph, all he/she sees is that the Per Capita Income is increasing at a fixed rate between 2000 and 2014 but that is due to altering the graph by using cumulative measure. Each bar represents the sum of that particular year and all the previous years Per Capita Income. But Actually, the Per Capita Income is almost same from 2000 to 2014.

4. Number of People below poverty level in a country from 2006 to 2017

Manipulation Technique Implemented: Cumulative Axes



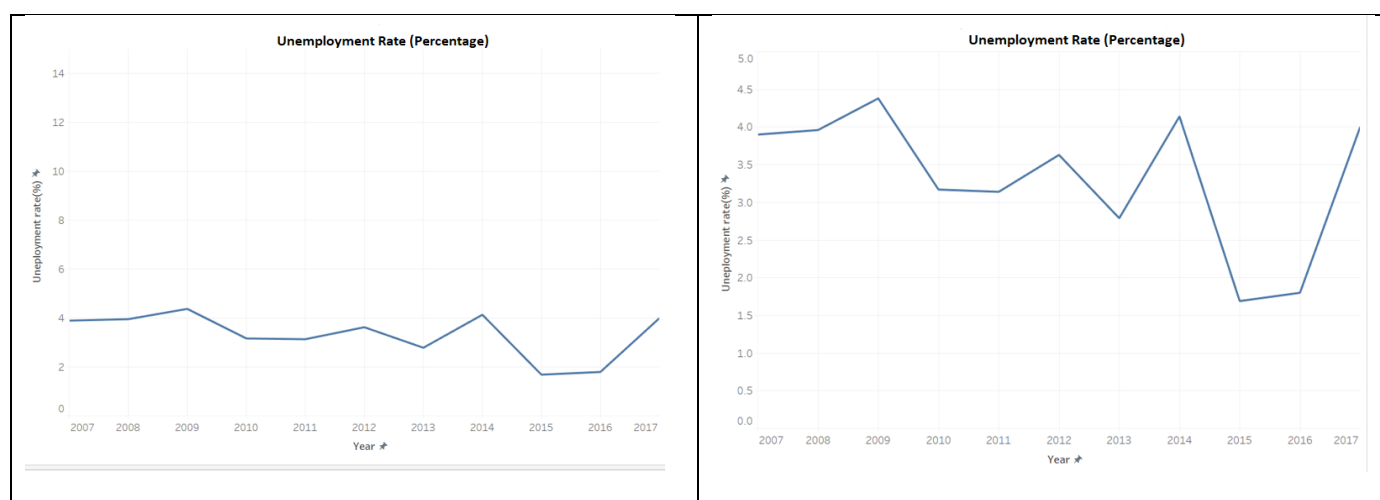
Manipulated Graph

Correct Graph

Justification: By looking at the first graph above, it is quite evident at first sight that the number of people below poverty line is constantly increasing but if someone looks closely, he/she can identify that there is an increment in number of people below poverty line every year due to addition of number of poor people of current year added with those of previous years. But, in true scenario, the number is fluctuating throughout the years 2006 to 2017.

5. Unemployment rate in a country from 2007 to 2017

Manipulation Technique Implemented: Compressed Axes



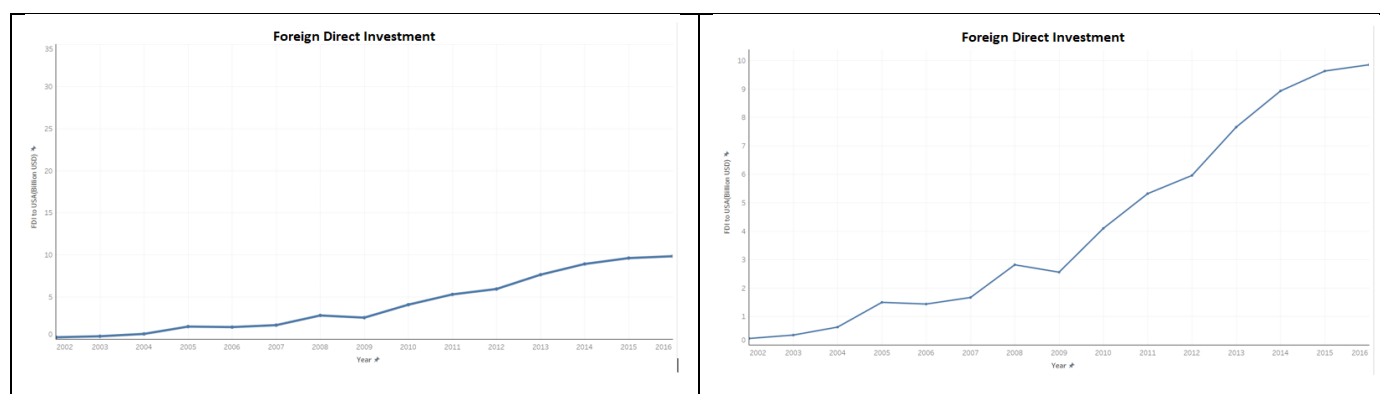
Manipulated Graph

Correct Graph

Justification: Unemployment is an important measure for any country. The first chart shows almost no fluctuations in Unemployment rate percentage due to compression of the Y-Axis. This gives a false image of stability to the participant. The true picture is the unemployment is varying frequently with each passing year and is evident from the second graph.

6. Foreign direct investment of a country from 2002 to 2016

Manipulation Technique Implemented: Compressed Axes



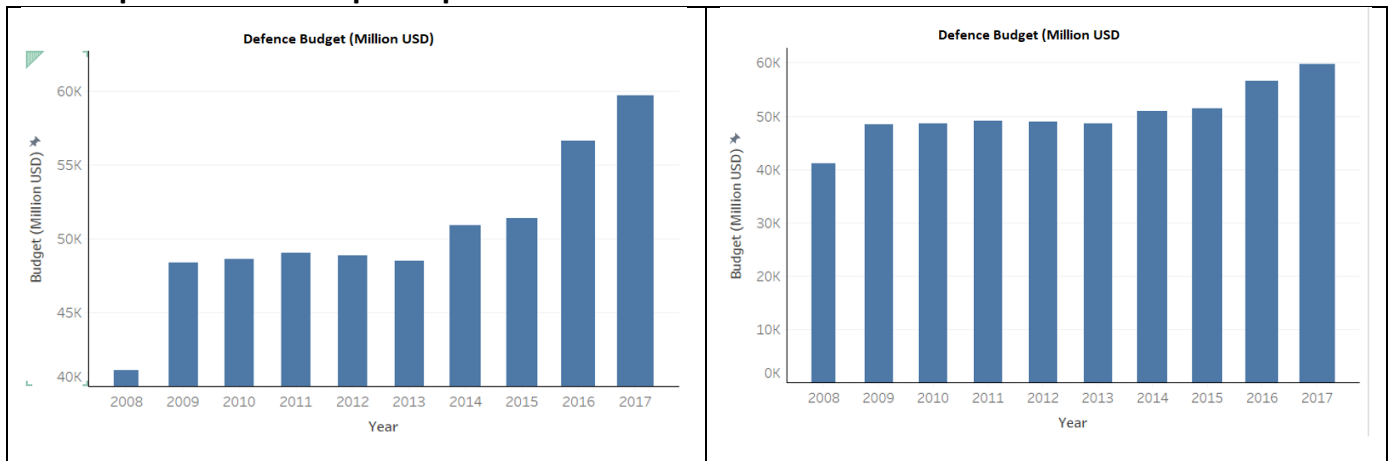
Manipulated Graph

Correct Graph

Justification: The above graph illustrates the variation of FDI from year 2002 to 2016. A participant can easily observe negligible increments in FDI by every passing year from the first graph. This is due to the fact the Y-axis has been compressed which has the interval of only 5 unit as compared to 1 of the second graph. But in fact, the FDI is highly fluctuating which can be observed by seeing the second graph.

7. Defence budget of a country from 2008 to 2017

Manipulation Technique Implemented: Truncated Axes



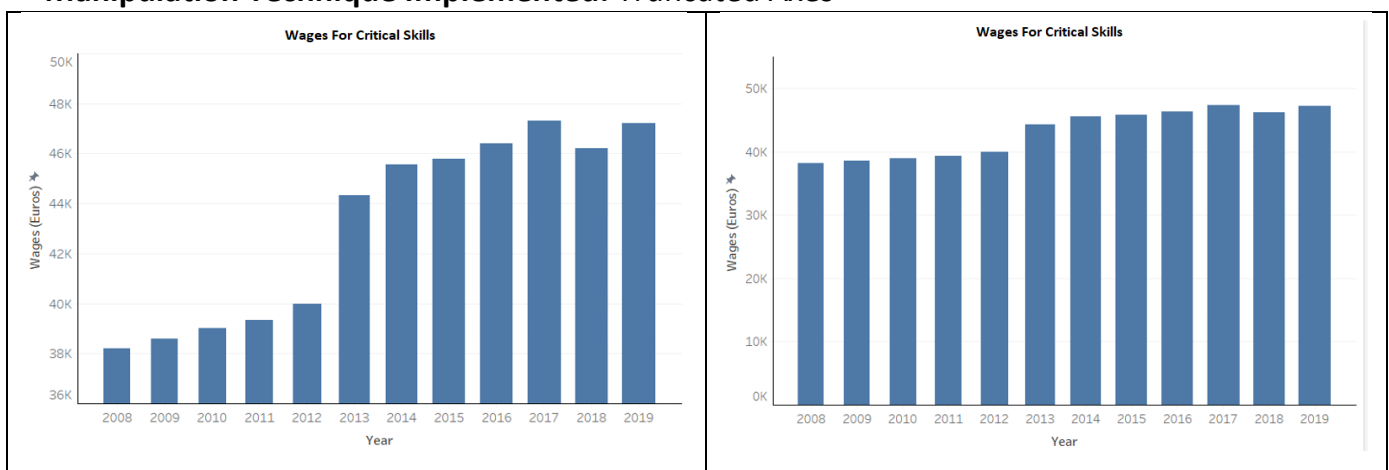
Manipulated Graph

Correct Graph

Justification: Many people have a habit of comparing the length of adjacent bars in order to understand the ratio between them. So, by truncation the Y-Axis, we can manipulate the participants. In first chart above, it appears that the defence budget of 2009 is almost 5 times than that 2008. But when we observe the correct chart, it becomes evident that there is only a mild increase in the defence budget from 2008 to 2009.

8. Wages for critical skills in a country from 2008 to 2019

Manipulation Technique Implemented: Truncated Axes



Manipulated Graph

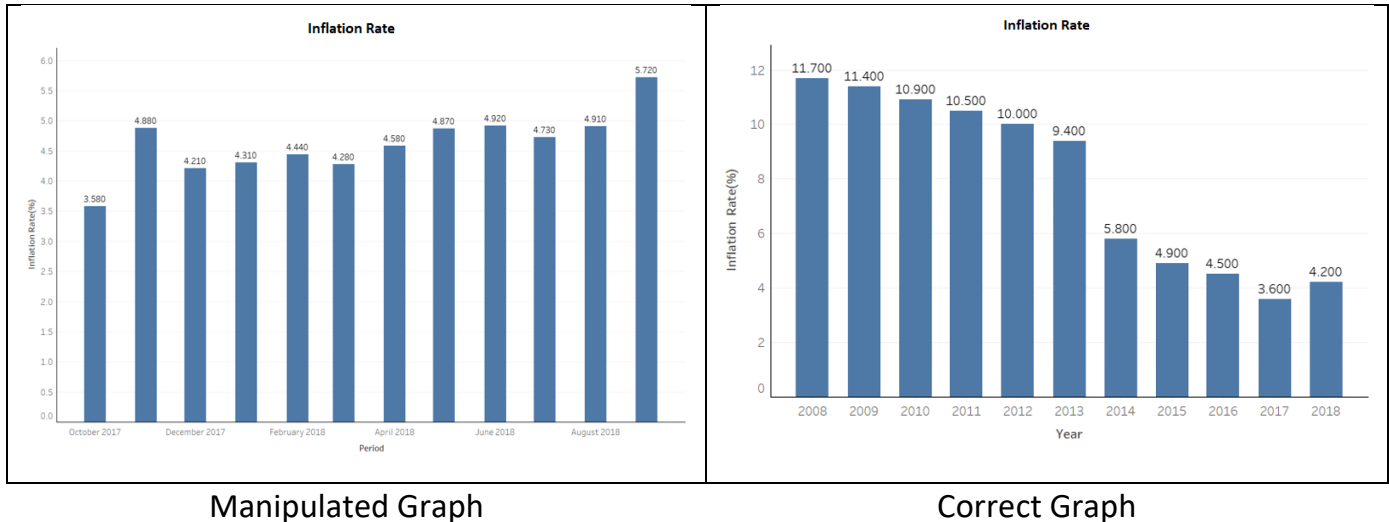
Correct Graph

Justification: The misleading chart above, depicts that wages for critical have doubled between the years 2012 and 2013. This has happened because of truncating

of Y-Axis. Actually, if we check the accurate representation, it can be concluded that it is only slightly increasing.

9. Variation of Inflation rate of a country

Manipulation Technique Implemented: Cherry Picking



Manipulated Graph

Correct Graph

Justification: The inaccurate representation of graph show increment in the inflation rate with slight fluctuations, but the correct one shows a fall in the inflation rate. This ambiguity is happening due to cherry picking of data in the first case where only the month of a particular year has been chosen instead of years.

3. Pilot Experiment

4.1 Data Analysis

The data was collected from 100 participants who volunteered to be the part of the experiment. Out of the 100 participants, 80 fastest responses were considered for analysis because we wanted the participants to provide their first impression of the graph.

Answering one manipulated and one correct graph which represent the same dataset will be considered as one instance. Our purpose is to find the number of instances where the opinion (option selection) of the participant has changed while answering the manipulated and the correct representation.

For each participant, there were a total of 9 pairs of responses (one pair being correct response for correct graph and an expected response from incorrect graph for same dataset). We calculated the ratio of our expected pair of responses against the total pair of responses. After calculating it for all 80 participants, an average was taken to find the Average opinion reversal rate.

Calculation of expected pair of responses to total responses for a candidate:

	Pair- 1	Pair- 2	Pair- 3	Pair- 4	Pair- 5	Pair- 6	Pair- 7	Pair- 8	Pair- 9
Manipulated Representation	1	1	0	1	1	1	1	1	1
Correct Representation	1	1	1	1	0	1	1	1	1
Opinion Reversal	1	1	0	1	0	1	1	1	1
	1 = Expected Response								
	0 = Random Response								

From the above table it can be seen that there has been a reversal of opinions for 7 set of pairs out of total 9 pairs. We can calculate percentage reversal of opinion for each candidate using the following formula :

$$\% \text{Reversal of Opinion} = \frac{\text{Observed Reversal of Opinion pairs}}{\text{Total set of pairs}} = \frac{7}{9} = 0.77 = 0.77 * 100 = 77$$

In this manner, we can calculate the %Reversal of opinion Rate for each candidate and take a mean of it on 80 selected candidates to get the average opinion reversal Rate .

$$\text{Average opinion Reversal rate} = \frac{\sum_{i=1}^{i=80} \% \text{ Reversal of opinion}}{\text{Total selected candidates}} = \frac{6097.6}{80} = 76.22$$

From the above Data Analysis, it was observed that there has been an opinion Reversal of 76.22%, which means that every 3 out of 4 individuals got their opinion reversed on average. Since this ratio is high, it validates our hypothesis that opinions can be reversed by using axis- manipulation techniques.

3.2 Reflections :

From the proposed pilot experiment, it was observed that :

Firstly, for the same dataset, a person's opinion changes when he is presented with two graphs, first being the manipulated and later being the correct one. Secondly while observing the graphs, the focus is shifted more towards the trends or pattern in the graph rather than the data represented on the axes. This mainly happens because of Gestalt Principle according to which, human mind has an innate disposition to perceive patterns in the stimulus based on certain rules i.e. trends and bars in the graph which seem to follow a pattern or show a sign of similarity between them.

The following were the learning outcomes from the proposed experimental design:

- 1) People generally tend to form a perception about the graph as soon as they glance on it.
- 2) Very often, people don't notice whether the graph starts with a zero or some pre-defined value i.e. they generally go with a gut feeling that a graph is bound to start from a zero scale.
- 3) The intervals on the axes are most habitually neglected in contrast to the trends in the graph. This was evidently seen when the y-axis was compressed for the unemployment and FDI datasets in the experiment.

- 4) While considering the Y-axis, it is a common presumption that the values will increase in bottom-top manner, which is not true in the case of inverted axis technique.
- 5) It is very easy to convince the audience that the graph is ever increasing by using cumulative-axis. As a result of this the axes labels are often neglected and thereby convey the improper trend.

We were able to understand that given limited time and presented with manipulated axes, it is quite easy to confuse the audience into believing the false trends. Also, majority of the people rely on their basic instincts for evaluating a graph rather than observing both the axes and the labels carefully. The moment they lay their eyes on a chart, they make a perception about it which makes them more prone to mistakes, which can be evidently seen from our conducted experiments.

Thus, it can be concluded that various parameters like FDI, Rate of unemployment and GDP which greatly influence the development index of the country, when represented using manipulated visualization techniques can easily reverse the opinions of the people. These minute alterations in data representation can have a devastating effect on the people in the country and may even result in the election of an undeserving party and vice versa.

5 References

- [1] <https://www.livescience.com/45083-misleading-gun-death-chart.html>
- [2] <http://passyworldofmathematics.com/misleading-graphs/>
- [3] https://en.wikipedia.org/wiki/Principles_of_grouping
- [4] <https://explorable.com/confounding-variables>
- [5] https://csmoodle.ucd.ie/moodle/pluginfile.php/100463/mod_resource/content/3/08%20-%20Evaluation.pdf
- [6] <https://www.statista.com/statistics/233138/number-of-people-living-below-the-poverty-in-the-us/>
- [7] <https://data.worldbank.org/country/india>

6 Survey Questionnaire for Participants

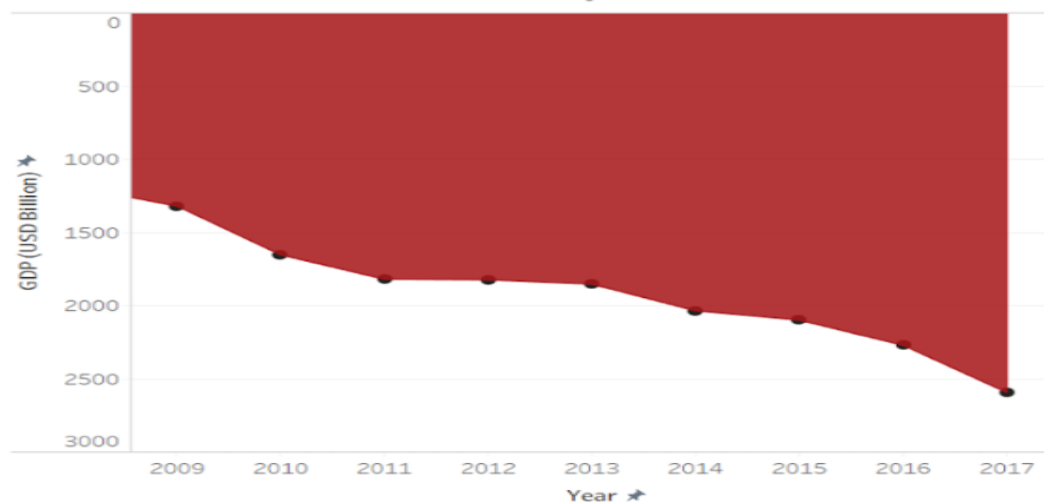
Opinion Reversal By Axis Manipulation

Rules for participants:
Complete the test within 6 minutes.
Attempt all the questions

SECTION -1

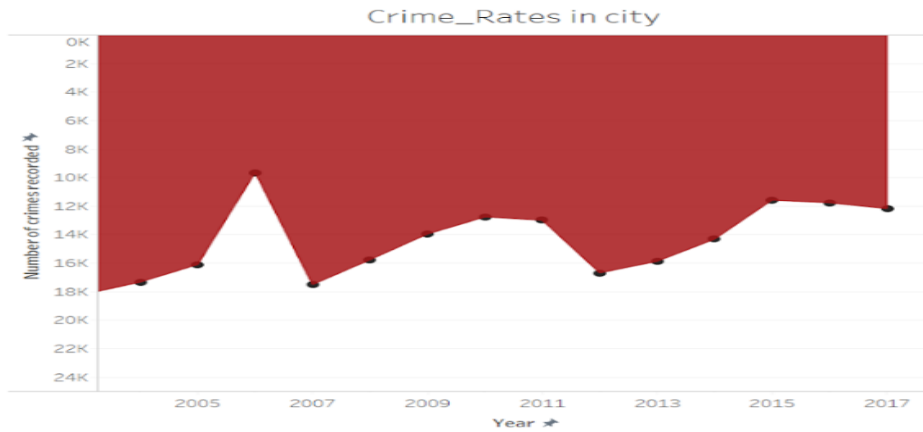
Q1-Please select the best possible option from the below graph

GDP of Country in USD Billion



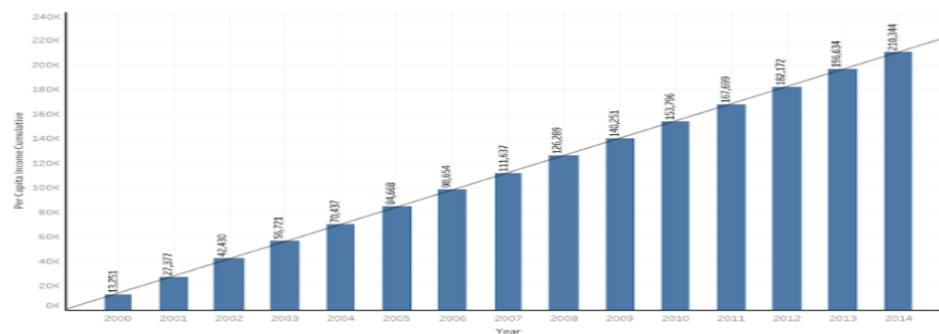
- ☐ The GDP of the country is consistently increasing
- ☐ The GDP of the Country is neither increasing nor decreasing
- ☐ The GDP of the country is fluctuating
- ☐ Don't know

Q2-Please select the best possible option from the below graph



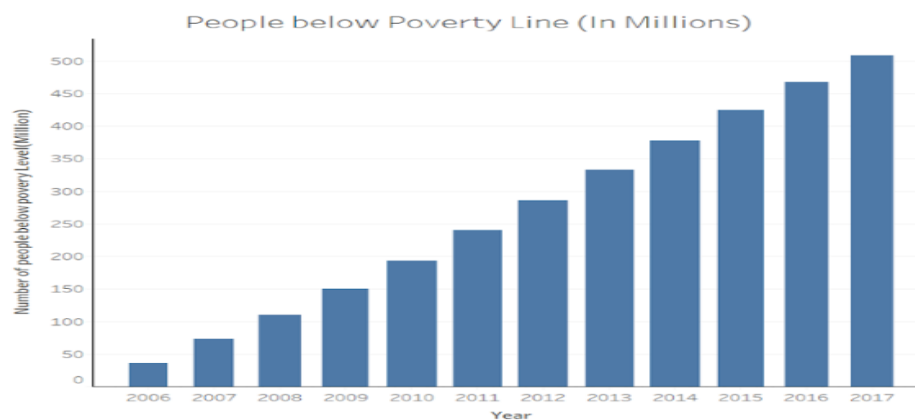
- ☐ Crime rate is fluctuating but increasing
- ☐ Crime rate is fluctuating but decreasing
- ☐ Crime rate is sharply decreasing from 2007 to 2010
- ☐ Crime rate is always increasing

Q3-Please select the best possible option from the below graph



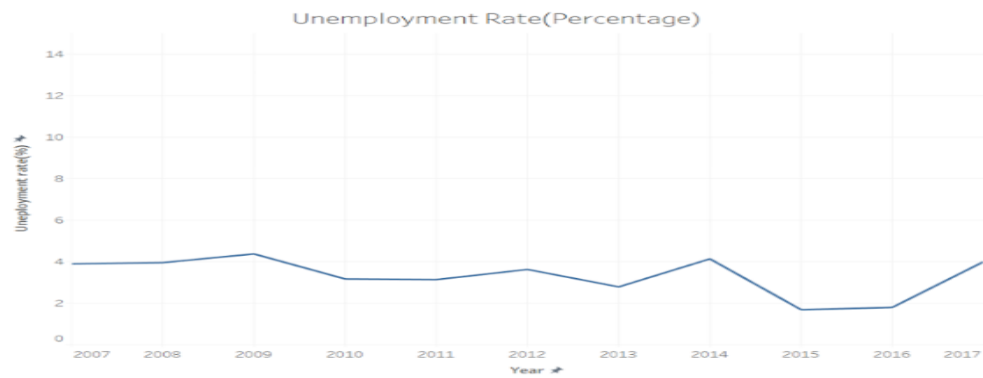
- ☐ Per Capita Income is stable
- ☐ Per Capita Income is continuously increasing
- ☐ Per Capita Income is continuously decreasing
- ☐ Don't Know

Q4-Please select the best possible option from the below graph



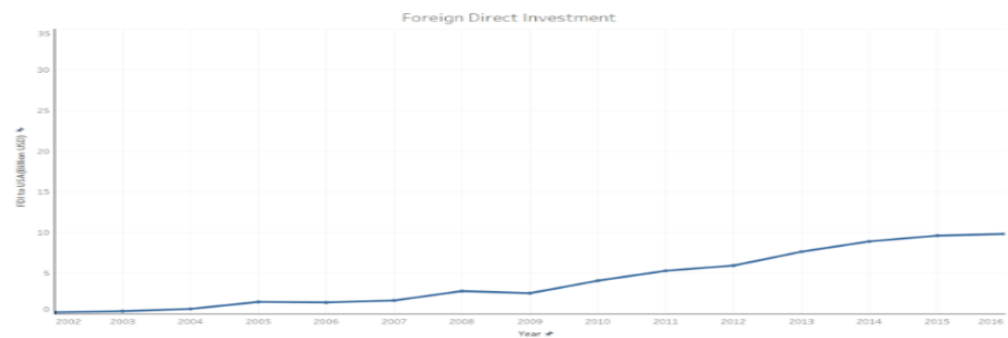
- ☐ Number of people below poverty line has decreased from 2007 to 2017
- ☐ Number of people below poverty line has increased from 2007 to 2017
- ☐ Number of people below poverty line is fluctuating
- ☐ Don't Know

Q5-Please select the best possible option from the below graph



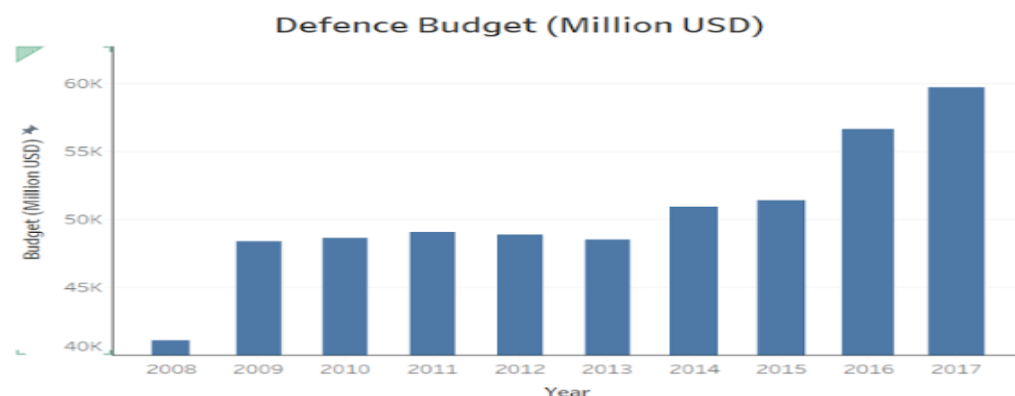
- ☐ The unemployment percentage is highly fluctuating
- ☐ The unemployment percentage is mildly fluctuating
- ☐ The unemployment percentage is strictly increasing
- ☐ Don't Know

Q6-Please select the best possible option from the below graph



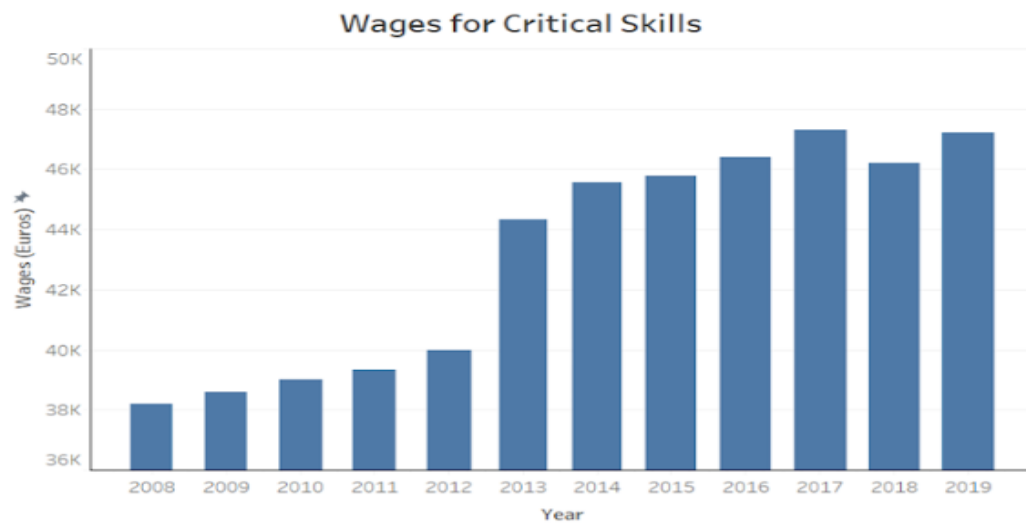
- ☐ The FDI is sharply increasing
- ☐ The FDI is mildly increasing
- ☐ The FDI is fluctuating but not increasing
- ☐ Don't know

Q7-Please select the best possible option from the below graph



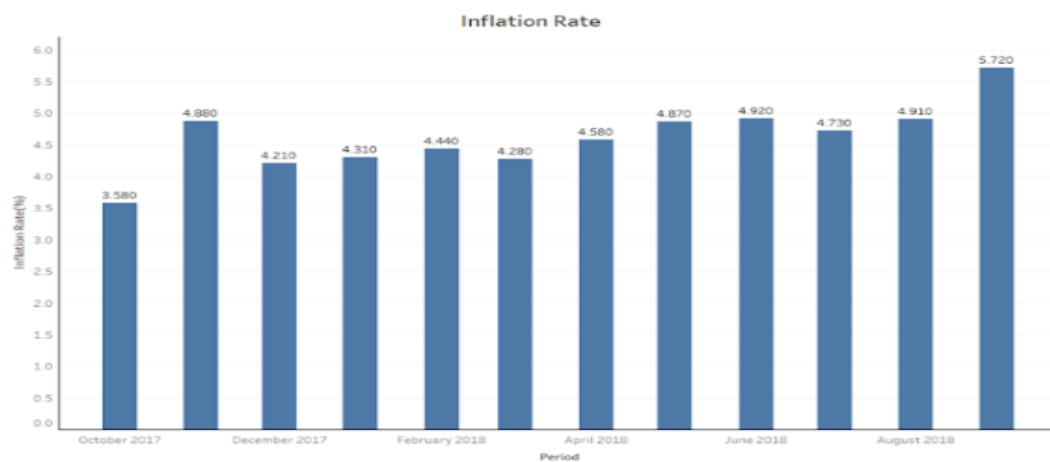
- ☐ The Defence budget has significantly increased from 2008 to 2009
- ☐ The Defence budget is continuously increasing
- ☐ The Defence budget is not increasing at all
- ☐ Don't know

Q8-Please select the best possible option from the below graph



- ☐ The Wages have significantly increased from 2012 to 2013
- ☐ The Wages are continuously increasing
- ☐ The Wages are Fluctuating
- ☐ Don't Know

Q9-Please select the best possible option from the below graph

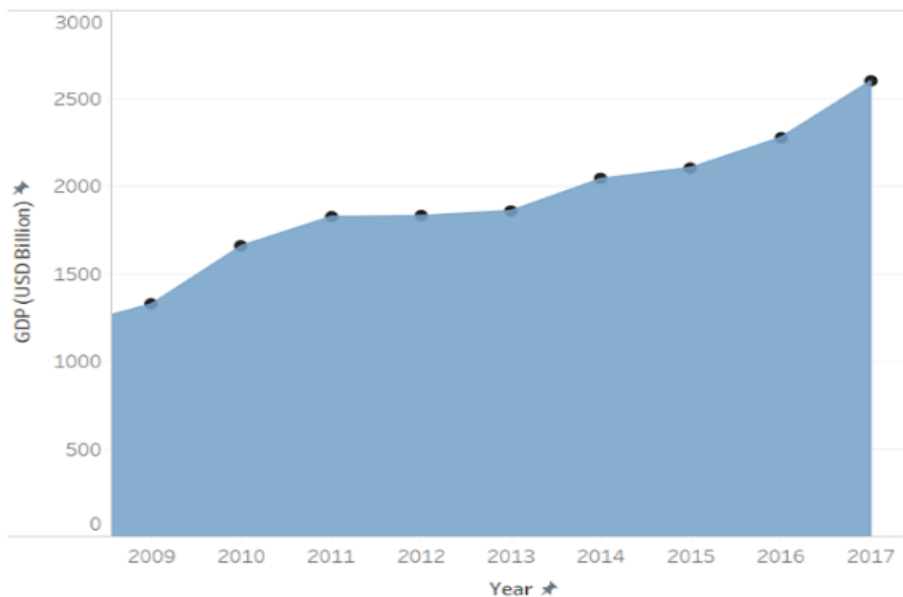


- ☐ The Inflation Rate is decreasing
- ☐ The Inflation Rate is increasing
- ☐ The Inflation Rate is neither increasing nor decreasing
- ☐ Don't know

SECTION - 2

Q1-Please select the best possible option from the below graph

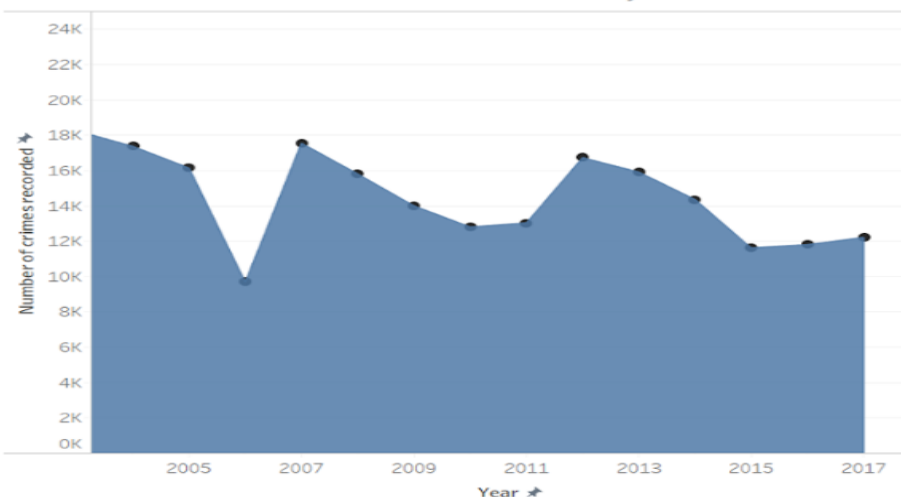
GDP of Country in USD Billion



- ☐ The GDP of the country is consistently increasing
- ☐ The GDP of the Country is neither increasing nor decreasing
- ☐ The GDP of the country is fluctuating
- ☐ Don't know

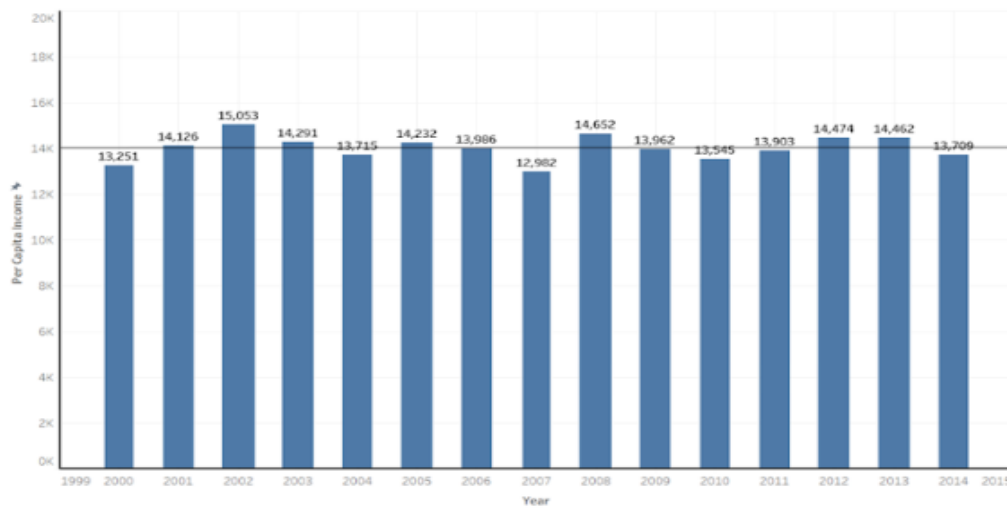
Q2-Please select the best possible option from the below graph

Crime_Rates in city



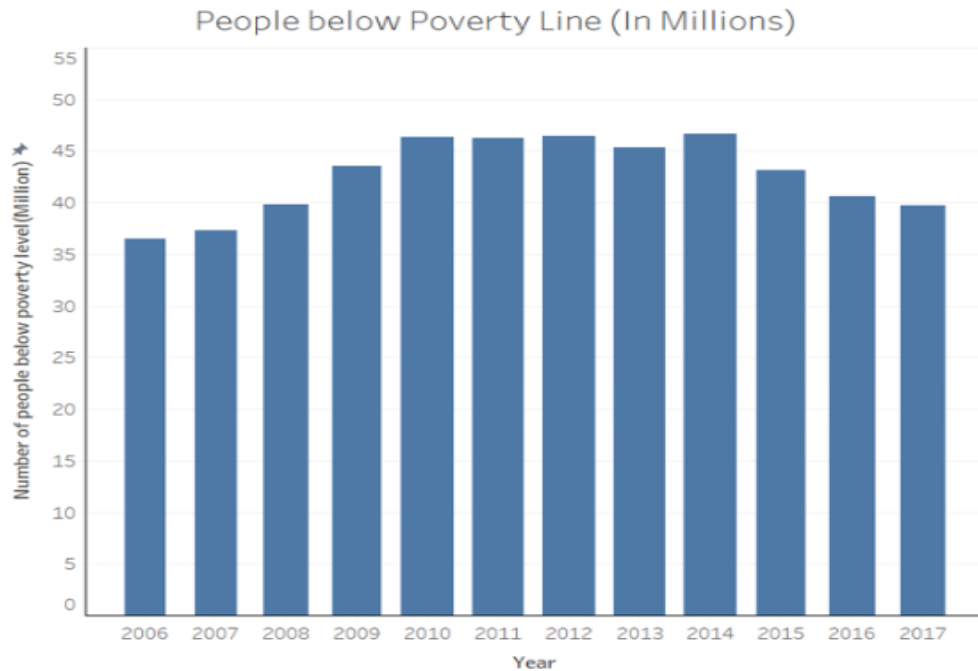
- ☐ Crime rate is fluctuating but increasing
- ☐ Crime rate is fluctuating but decreasing
- ☐ Crime rate is sharply decreasing from 2007 to 2010
- ☐ Crime rate is always increasing

Q3-Please select the best possible option from the below graph



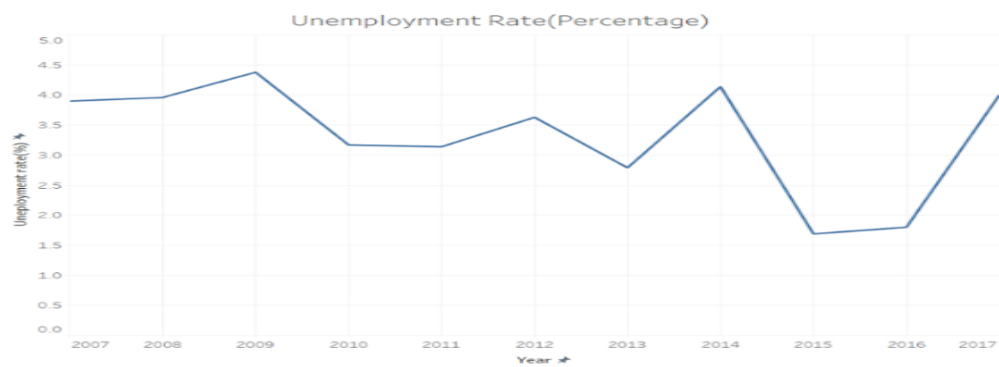
- ☐ Per Capita Income is stable
- ☐ Per Capita Income is continuously increasing
- ☐ Per Capita Income is continuously decreasing
- ☐ Don't Know

Q4-Please select the best possible option from the below graph



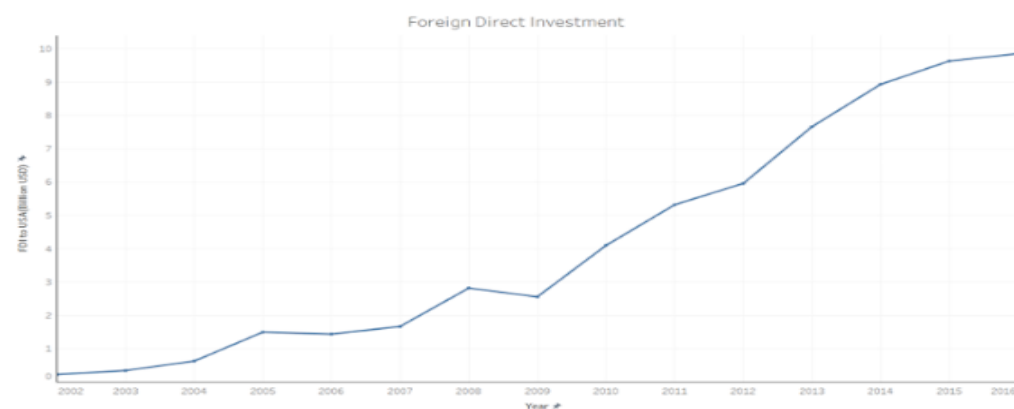
- ☐ Number of people below poverty line has decreased from 2007 to 2017
- ☐ Number of people below poverty line has increased from 2007 to 2017
- ☐ Number of people below poverty line is fluctuating
- ☐ Don't Know

Q5-Please select the best possible option from the below graph



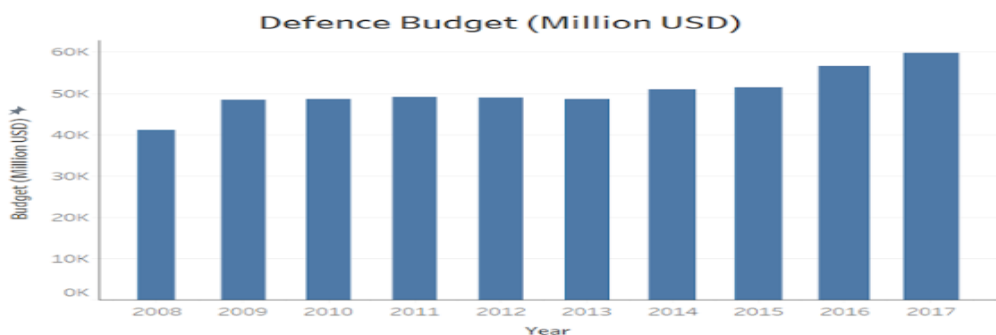
- ☐ The unemployment percentage is highly fluctuating
- ☐ The unemployment percentage is mildly fluctuating
- ☐ The unemployment percentage is strictly increasing
- ☐ Don't know

Q6-Please select the best possible option from the below graph



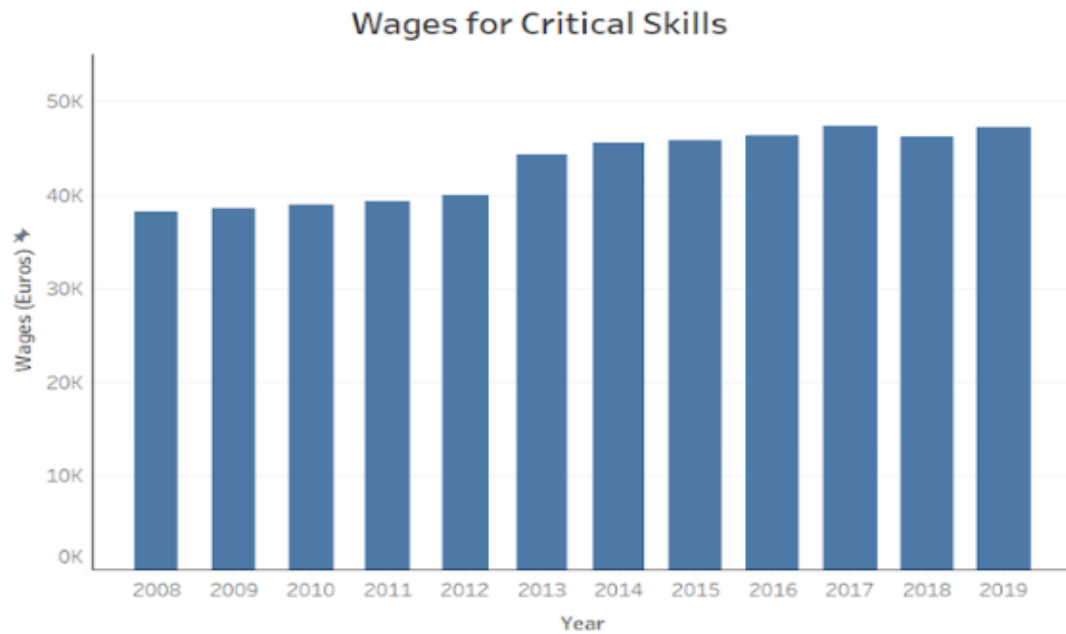
- ☐ The FDI is sharply increasing
- ☐ The FDI is mildly increasing
- ☐ The FDI is fluctuating but not increasing

Q7-Please select the best possible option from the below graph



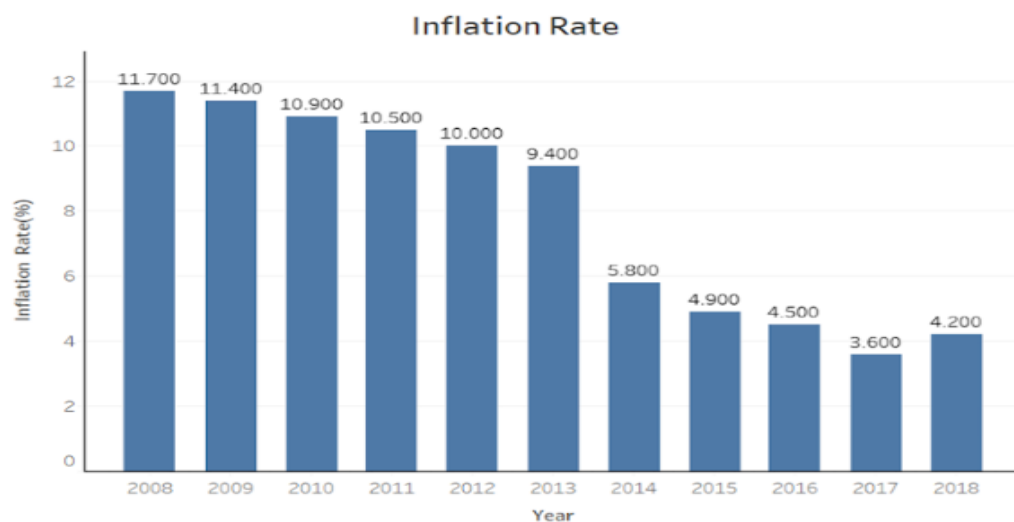
- ☐ The Defence budget has significantly increased from 2008 to 2009
- ☐ The Defence budget is continuously increasing
- ☐ The Defence budget is not increasing at all
- ☐ Don't Know

Q8-Please select the best possible option from the below graph



- ☐ The Wages have exponentially increased from 2012 to 2013
- ☐ The Wages are continuously increasing
- ☐ The Wages are Fluctuating
- ☐ Don't know

Q9-Please select the best possible option from the below graph



- ☐ The Inflation Rate is decreasing
- ☐ The Inflation Rate is increasing
- ☐ The Inflation Rate is neither increasing nor decreasing
- ☐ Don't know