Homework 1

Statistical Inference

Abbas Nosrat
810199294

1. :
    a) :
        - The conductor only examined patients and no random assignement was performed. Therefore, this case is an observational study.
        - The exolanatory variable is having heart disease and the response variable is having PTSD.
        - This study is observational and cannot establish any causal relationship
    b) :
        - Due to random asignement, this study is experimantal.
        - The explanatory variable is the diet type and the response variable is the amount of weight loss.
        - This study is experimental Therefore it can establish a causal relationship.
    c) :
        - Random assignement has taken place among two groups therefore the study is experimental.
        - The explanatory vriable is the education method and the response variable is the performance of students at the exame.
        - This study is experimental Therefore it can establish a causal relationship.

2. :
    a) The cofounding variable is time or the seasonal change to be precise. As the weather gets warmer, the demand for icecream increases. It can also be infered that murderers tend to commit more homicide when the weather gets warmer. However the demand for icecream and the number of murders have no causation.
    b) There is no cofounding variable in this case. Cancer and proximity from a high voltage post have a causal relationship. There is a phenomena called Crona. When the voltage is high enough, the cables shoot radical electrons and those electrons cause cancer.
3. :
    a) This is a cluster sampling method. The religious meetings are clusters. The conductors have surveied the entire population of each randomly selected religious group.

b) This case is most simular to the multi stage sampling method. At the first stage a random starting point is selected and at the second stage, the samples are selected systematicly.
c) This case is simple random sampling. The samples can be devided in groups. However it is not stated that the devision occurred before sampling. Regardless of the sampling method, samples can be devided into groups after sampling.
d) Simple random sampling. The samples are randomly selected.

4. :

a) First of all, the subjects must be devided in groups based on their level of experties. Second, each group must be devided in two equal groups and assigned to Whatsapp and Telegram. The devision and assignement must take place randomly. Finaly, the subjects are asked to perform various tasks and performance time must be recorded.
b) Yes blocking method was utilized. There was a variance in the level of experty for each test subject and it affected the test results and created a bias in the conclusions. Therefore blocking was utilized.

5. :

a) Not all the passengers had a cellphone and the answer for that group of the passengers may have been deifferent. Among the passengers who had cellphones, some of them did not answer their phones. In other words, the study was volentary and the answer of those who didn't take part in that study may have been NO. Hence the results are biased.
b) Only the students enrolled in the class have answered the test but the answer may be different outside of the class. Hence the results are biased.
c) The sampling method is completely random the subjects didn't know the lable(blind testing was used) . Therefore the results are unbiased.
d) The sampling was occurred at a specific temporal point. In other words, had the survey been conducted another day, the number of sexual abuse victums could be more.Furthermore, the survey question is a personal one and people are not likely to answer honestly. Therefore the results are biased.
e) The first source of bias is the envoirment. The people not using instagram may have a strong unsatisfaction with snapp. The second source of biase is the reward. If it is stated that a reward is given at the end of this survey, people will tend to have a positive answer.

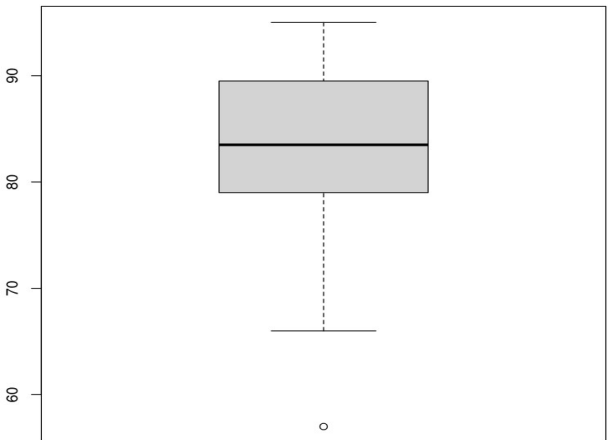
6. :

a) False. The plot skipped a few months between each data point and there could be a decrease or a spike in income.
b) False. The population of cities is not specified. Blue cities may have more population than the red cities.
c) False. The decrase is not strong because the slope is not that sharp and it is not monotonous. Meaning that there are some increases along the way.

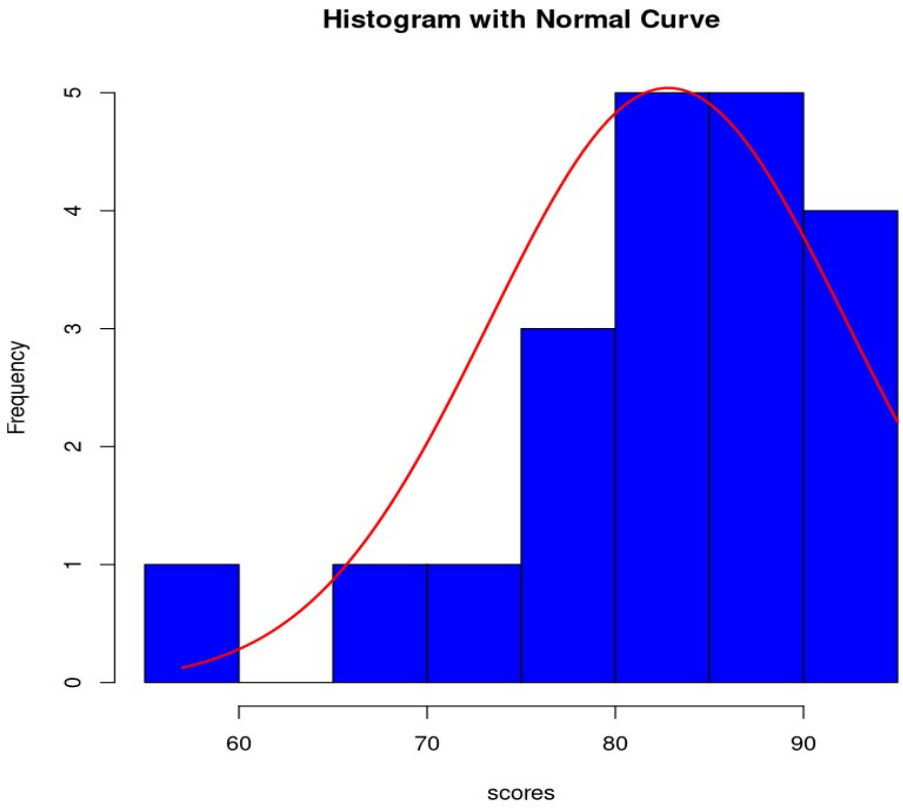
d) False. The plot is only for a few months and climate change is natural. To give a statement regarding global warming, one must plot the average tempreture over the course of a few years.

7. :
   a) H_0 = Average MUAC of smokers is 24cm or smaller. H_A = Average MUAC of smokers is larger than 24cm.
   b) Based on the results, only 2 experiments had a vkue of 24 or less. Therefore, the p-value is1- 0.04 which is 0.96.
   c) This means that Rasul is 96% confident that the researchers claim is false.
8. :
   a) c() command was used
   b) :
      - mean: 82.8
      - median: 83.5
      - var: 90.1684210526316
      - SD: 9.49570540047613
      - mode: 79 (mode had no builtin function so it was implemented from scratch)
   c) quantile function was used to find 0.025 and 0.975 percentiles and anyting below and above those values is treated as outliers.
      - Upper bound =94.525
      - lower bound=61.275
      - outliers = 57,95

   d) :
      -



e)



i. The data is right skewed
ii. since the frequency is larger for values above the median, the mean is expected to be higher. In other words larger than median values have more frequency.
iii. Median. Because distribution is not normal and it is right skewed. This affects the mean value and drags it towards the higher end of the data whereas median is always the center point of the dataset.

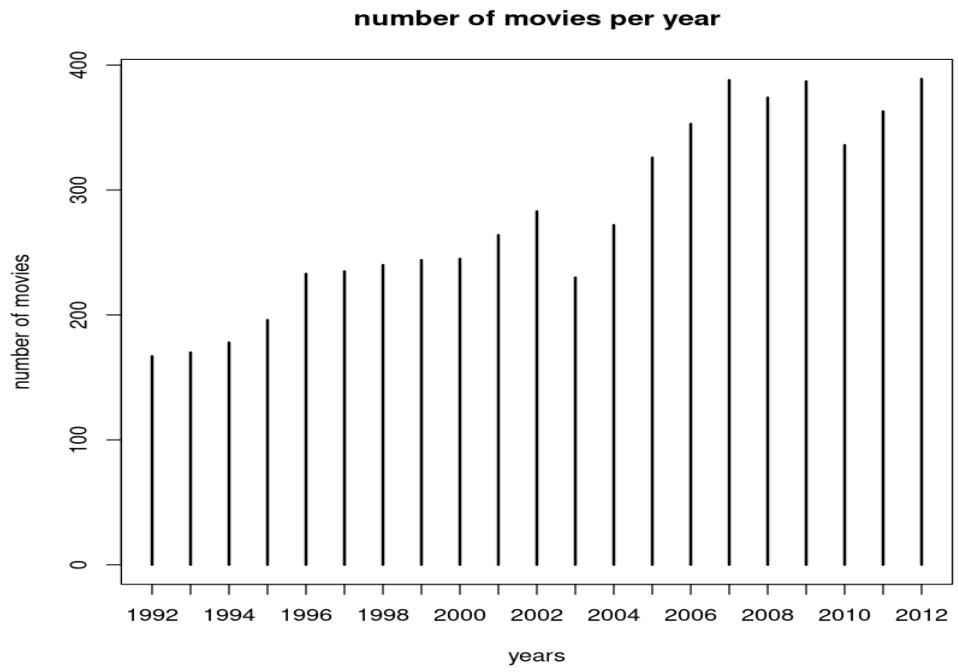
9. :

(a)

title	year	duration	total_votes	budget	USA_gross_income	worldwide_gross_income	tomatometer_status	audience_rating
index	cat	num	num	num	num	num	cat	num

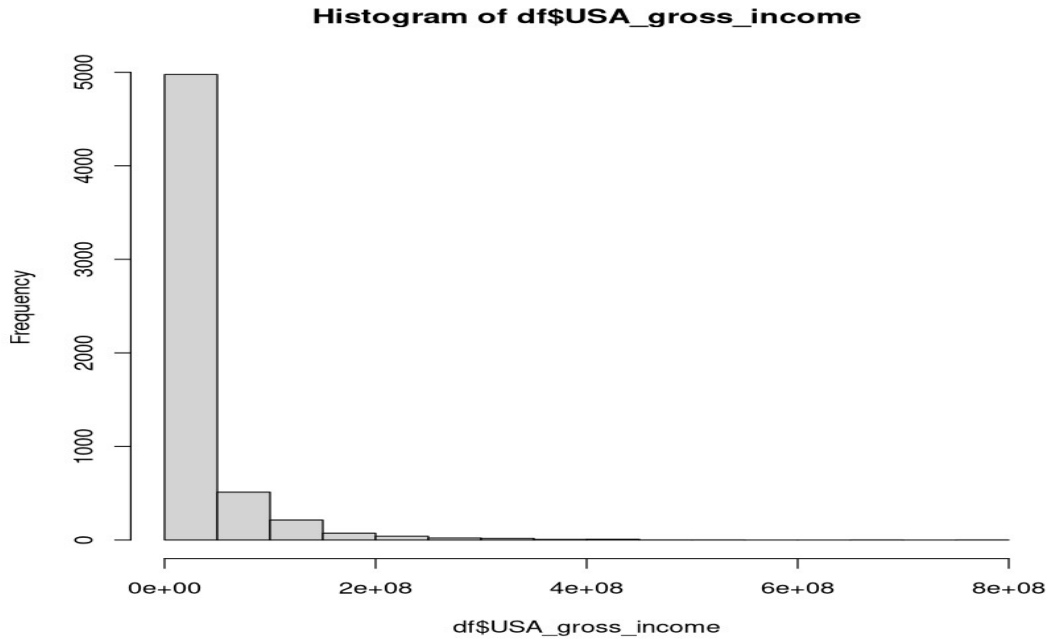
Title is called index because it is not numeric and it is unique to each entry.

(b)



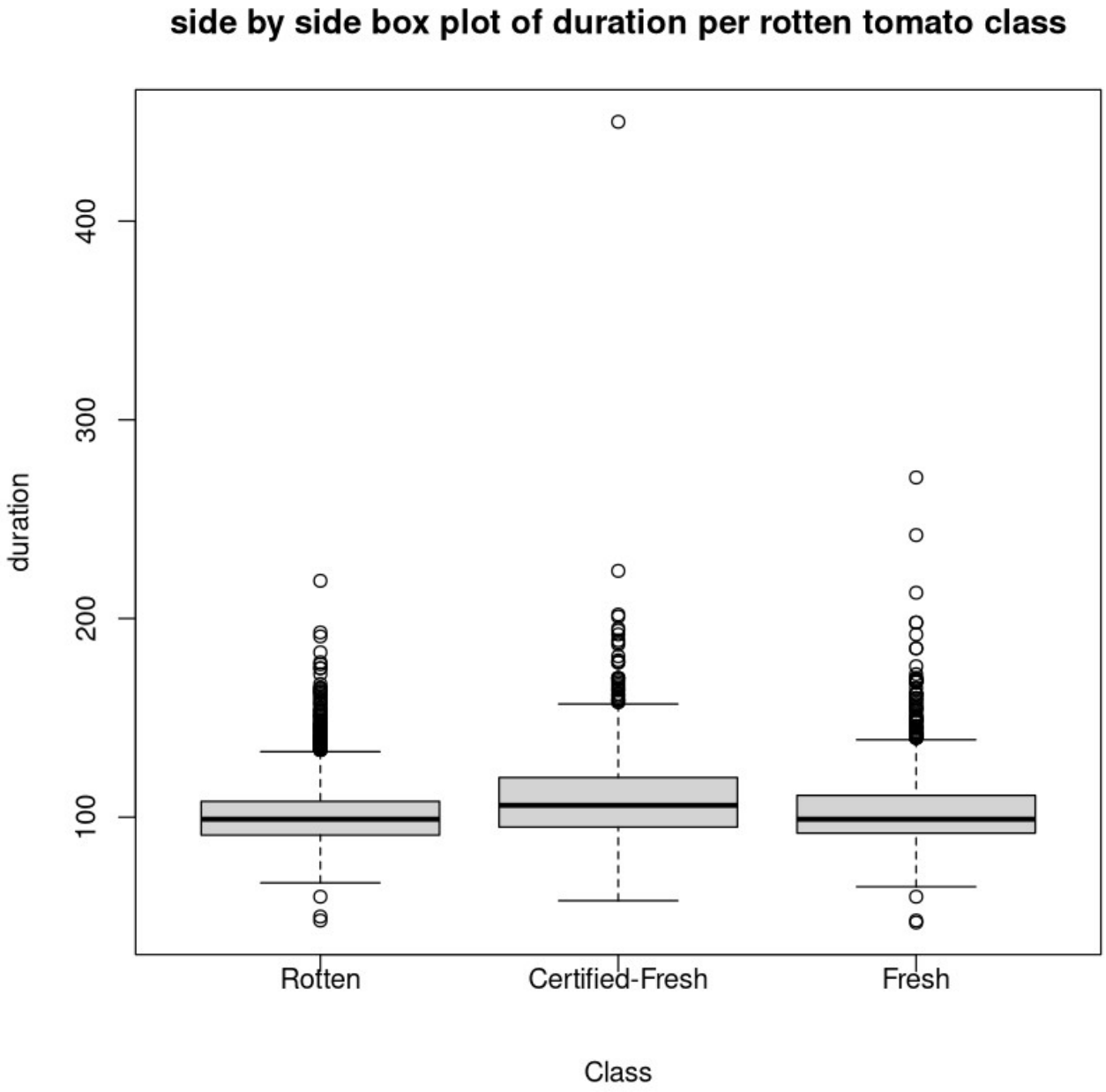
Bar plot is appropriate for comparison between years.

(c)



The data is right skewed. Which means that most movies have less income than the median value.

(d)

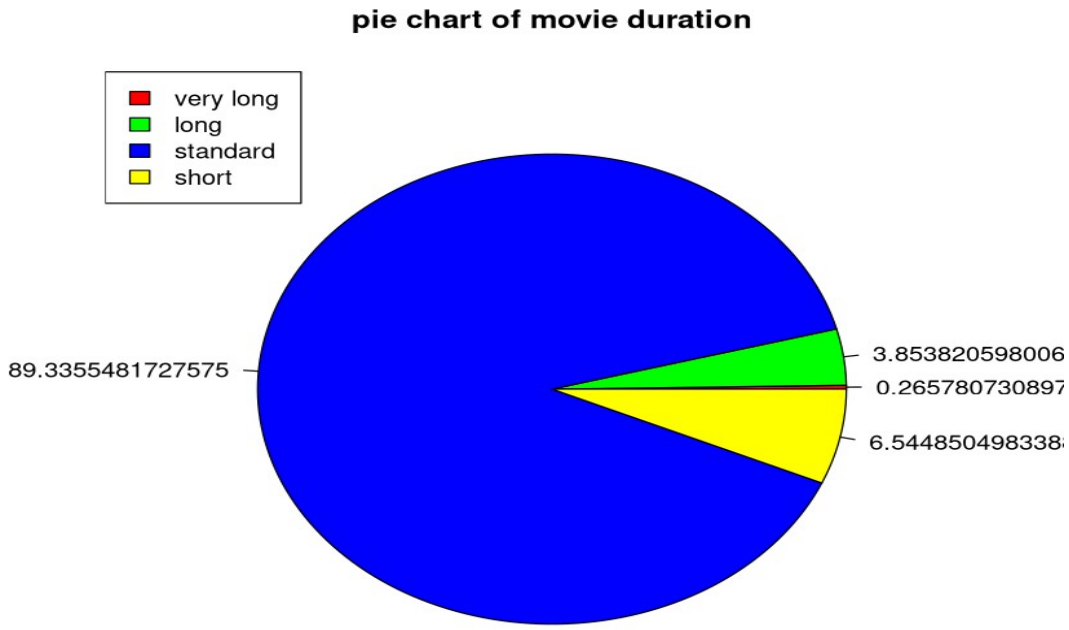


In all classestliers are above the rest of the data exept “ Fresh” class which has a few outliers under the box. The number of outliers are far more than the rest in “ Rotten” case.

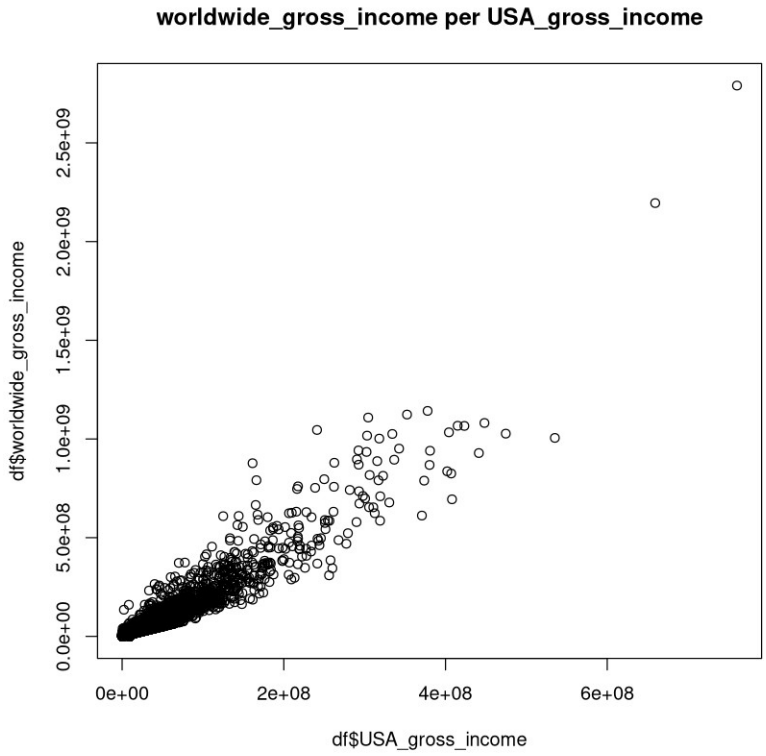
```
[1] "Rotten"

 144 · 165 · 48 · 74 · 140 · 191 · 76 · 148 · 183 · 137 · 75 · 138 · 153 · 140 · 154 · 158 · 76 · 164 · 157 · 137 · 175 · 145 ·
 162 · 151 · 67 · 75 · 165 · 163 · 147 · 78 · 76 · 79 · 172 · 159 · 79 · 78 · 79 · 151 · 76 · 138 · 137 · 74 · 143 · 74 · 149 · 77 ·
 77 · 72 · 142 · 147 · 75 · 79 · 77 · 77 · 79 · 78 · 140 · 140 · 73 · 137 · 175 · 79 · 219 · 167 · 78 · 76 · 140 · 76 · 145 · 78 ·
 137 · 150 · 139 · 138 · 72 · 193 · 157 · 144 · 76 · 139 · 76 · 152 · 146 · 139 · 161 · 178 · 145 · 155 · 146 · 78 · 160 · 79 · 74 ·
 71 · 74 · 75 · 70 · 74 · 79 · 141 · 147 · 143 · 151 · 137 · 177 · 73 · 163 · 165 · 72 · 140 · 142 · 79 · 145 · 146 · 78 · 73 · 60 ·
 145 · 146 · 142 · 151 · 153 · 75 · 76 · 71 · 75 · 141 · 73 · 79 · 76 · 78 · 139 · 140 · 72 · 78 · 154 · 149 · 144 · 78 · 163 · 75 ·
 141 · 72 · 50 · 71

[1] "Certified-Fresh"

 78 · 202 · 77 · 178 · 178 · 170 · 170 · 78 · 73 · 78 · 170 · 58 · 166 · 78 · 165 · 78 · 75 · 167 · 189 · 187 · 224 · 168 · 188 ·
 75 · 450 · 164 · 73 · 76 · 192 · 76 · 59 · 75 · 68 · 75 · 169 · 195 · 77 · 71 · 77 · 181 · 78 · 73 · 71 · 63 · 164 · 178 · 201 ·
 179 · 194 · 78 · 60 · 63 · 170

[1] "Fresh"

 75 · 48 · 271 · 170 · 185 · 75 · 168 · 158 · 75 · 76 · 72 · 157 · 72 · 176 · 152 · 151 · 76 · 71 · 155 · 170 · 75 · 172 · 154 ·
 72 · 74 · 69 · 75 · 75 · 160 · 70 · 242 · 169 · 169 · 160 · 198 · 168 · 213 · 75 · 75 · 185 · 154 · 156 · 60 · 165 · 162 · 168 ·
 76 · 159 · 162 · 75 · 75 · 76 · 74 · 76 · 169 · 198 · 76 · 72 · 47 · 192 · 162 · 65 · 76 · 162 · 76 · 155
```

The method used in question 8 is used here to find the outlier values. Rotten class has the most outliers followed by fresh and certified fresh

e.



f.



There is a linear relation between the two variables.