Unethical Data Visualization

Lohit Marla

2023-11-11

Unethical Data Visualization: It is an unusual form of representing the data in the form of visualization where the data in the charts misleads the actual information or by favoring a specific organization or by degrading competitors. There are multiple forms we need to take into consideration while developing the charts omitting them which results in the unethical data visualization issues and are

1. Ensuring data accuracy and completeness
2. Respecting data privacy and consent
3. Clarity and simplicity
4. Fairness and objectivity
5. cultural sensitivity and inclusive

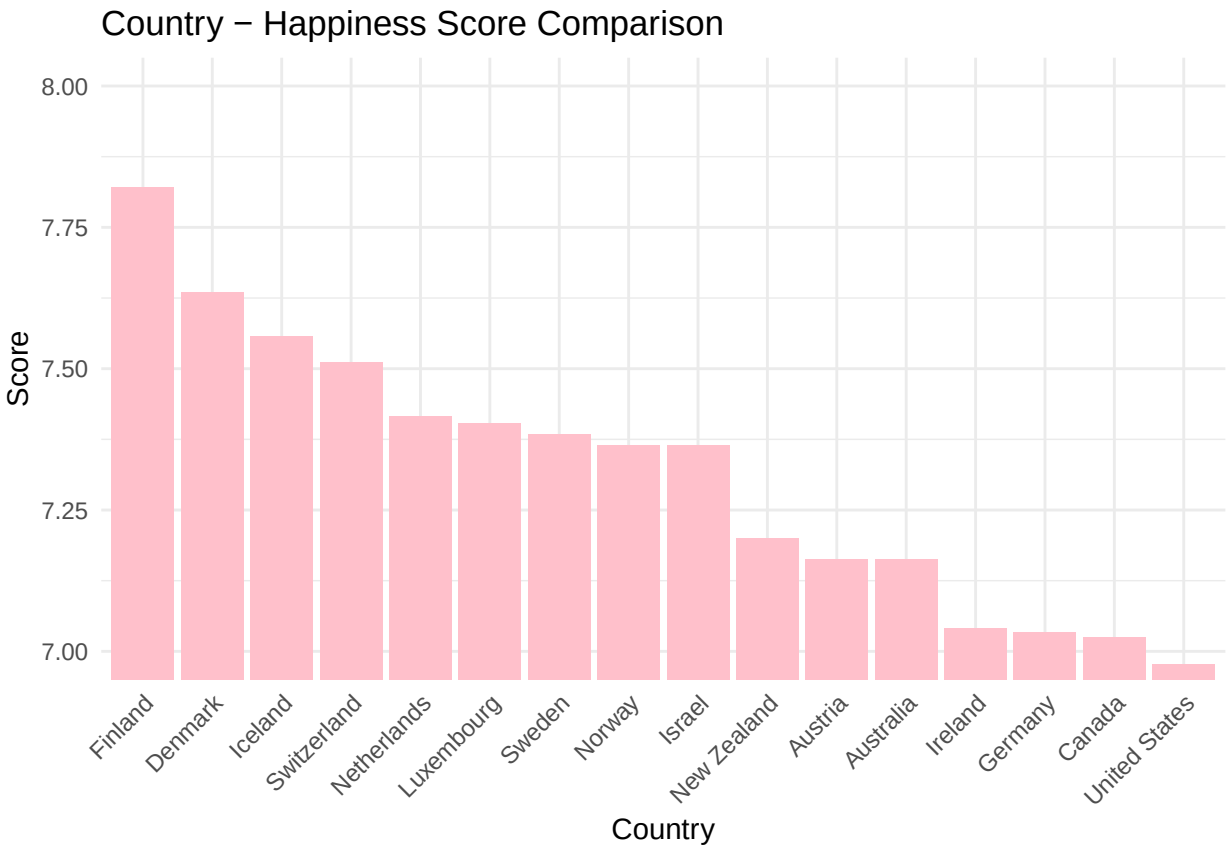
Following are the some forms of this immoral act

1. Truncating Y-axis

2. Reordering the data
3. Omitting Data
4. Correlating Causation
5. Cherry Picking

Following is one of the example to illustrate the immoral data visualisation

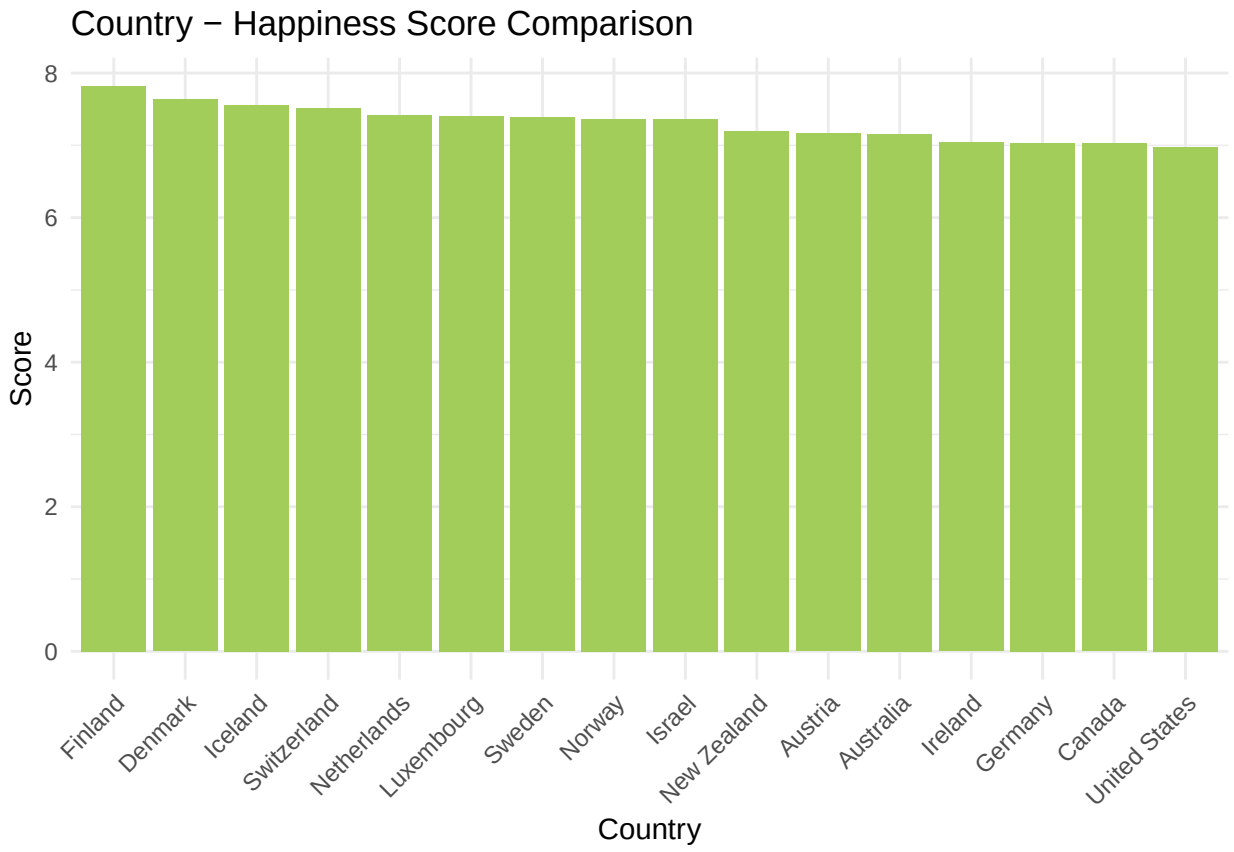
```
library("ggplot2")

happines_report <- read.csv('hapiness_report_2022.csv', header=TRUE)

happines_report <- head(happines_report, 16)

ggplot(happines_report, aes(x = reorder(country, -score), y = score)) +
  geom_bar(stat = "identity", fill = "pink") +  # Use geom_bar for bar chart
  labs(title = "Country - Happiness Score Comparison",
       x = "Country",
       y = "Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
coord_cartesian(ylim = c(7, 8))
```



The happiness report of 16 countries where the data represents the countries along with their happiness score. From the above bar chart we can observe there is a huge difference between the happiness score of United States with respect to the Finland. When if the same was published in the newspapers most of the public when goes thorough the chart interpret in such a way that the happiness index of US is too low and that country has many issues to deal with. This is all because of the truncating the data on the y-axis whereas in reality there is not much difference between the scores between the two countries scores as the actual difference is less than 1.

Instead the data visualization can be made more accurate by displaying the complete data of the y-axis without truncating which means from 0-8. Following can be done to improvise the code.

```
ggplot(happines_report, aes(x = reorder(country, -score), y = score)) +
  geom_bar(stat = "identity", fill = "darkolivegreen3") +  # Use geom_bar for bar chart
  labs(title = "Country - Happiness Score Comparison",
       x = "Country",
       y = "Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Country − Happiness Score Comparison
Score
8
6
4
2
0
Finland
Denmark
Iceland
Switzerland
Netherlands
Luxembourg
Sweden
Norway
Israel
New Zealand
Austria
Australia
Ireland
Germany
Canada
United States
Country