1 Visualization Analysis

1.1 New York City Weather in 1980 by Edward Tufte

Tufte's weather chart consists of three main subplots. The bottom subplot shows relative humidity as of noon as a percentage, which is a variable, presumably consisting of 365 samples in the dataset. The day of the year $1, 2, ..., 365$ is encoded with x-position, and the precipitation percentage with y-position. There are vertical gridlines indicating the start of each month, and horizontal gridlines for the quartiles.

The middle subplot presents the total precipitation for each month in comparison to the "normal" for that month. This is a continuous variable measured in inches, and is encoded as the length of the bar in the bar charts (equivalently the area of the bar is proportional to the preciptation). The categories "actual" and "normal" are encoded in two ways, texture and position. The "actual" 1980 preciptation is represented with solid black bars positioned to the left of the respective "normal" precipitation, whose texture is diagonal hatching. The total precipitation for 1980 is presented textually along with the normal for the whole year. The month in question is encoded by position, with January to the left of February to the left of March etc.

The top subplot presents temperature data. There are three time-series line plots, one for the actual temperature, one for the "normal" low temperature for a given day of the year, and another for the "normal" high temperature for a given day. Finally there is a small bar chart comparing comparing the 1980 annual temperature to normal annual temperature embedded in some free space of the temperature chart. It uses a similar format to the precipitation chart (1980 on the left in solid black, normal on the right in diagonal hatching).

Some of the tasks that Tufte's vis. facilitates are as follows:

- compare 1980 precipitation/temperatures to nomal
- compare temperatures/precipitation of each month to each other month in 1980 and normally
- identify warmest, coolest, most humid, most rainy months in 1980 and normally
- identify outlier temperatures, precipitation rates
- identify periodicity or lack thereof

2 Visualization Design

The dataset I have chosen consists of captured network packets in an Internet of Things (IoT) laboratory. The dataset is called CICIoT2023 (TODO CITE) and was created by simulating various kinds of cyber attacks on IoT devices using other IoT devices. The dataset is available in two forms; firstly the raw packet capture (`pcap`)

data for each type of cyber attacks as well as for simulated benign traffic; secondly an engineered subset of the data in csv format which was designed specifically to be used for training machine learning models to detect cyber attacks.

The code used to extract the features in the csv from the pcap is also available with the dataset.