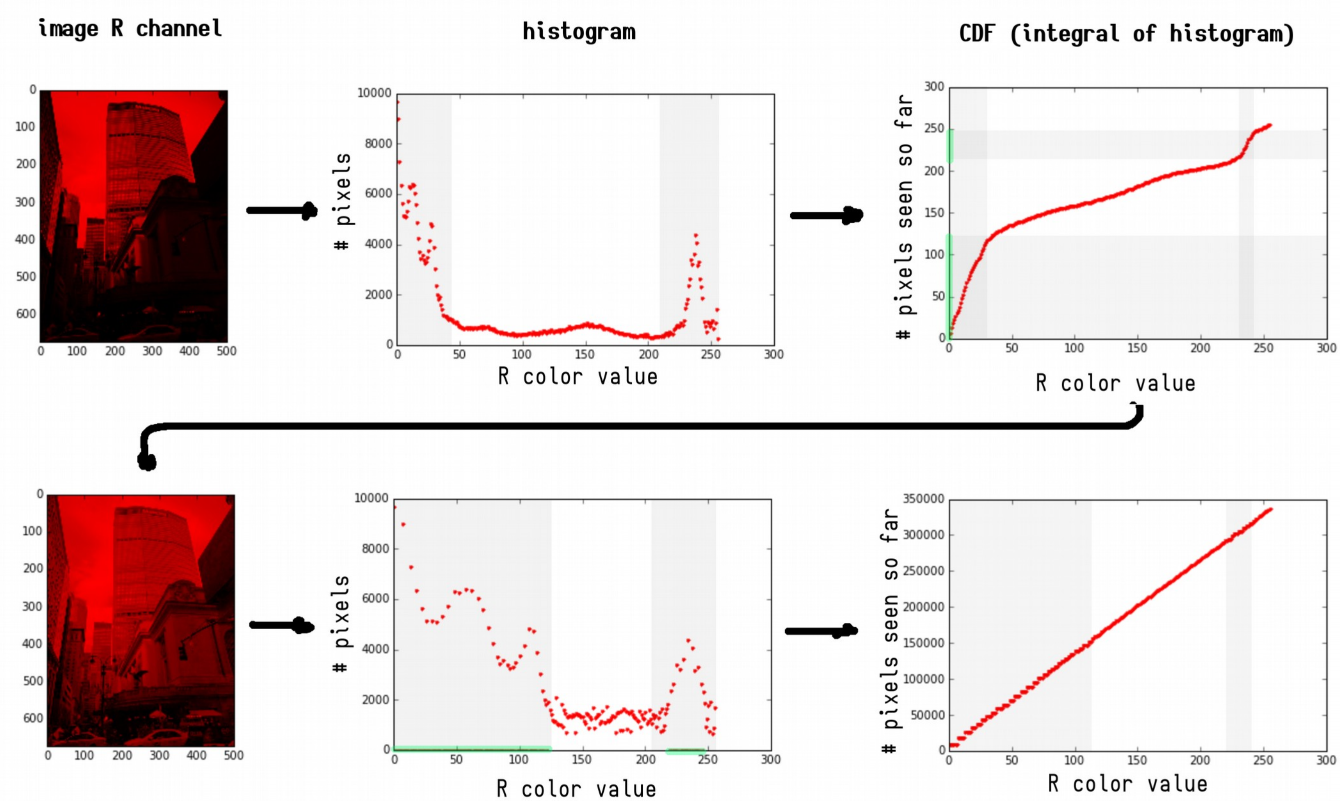
Quirks of Discrete Histogram Equalization

In trying to equalize this badly lit photo I took in NYC, we encountered some unusual behavior, which was initially thought to be a bug in the implementation. After a little analysis, it turns out that the issue was with the fact that the data we are working with is discrete, rather than continuous as the textbook would have us believe.

Below is an illustration of the process the program goes through to equalize an image. For simplicity, only the red channel has been included, but the process is the same for all channels.



The top row relates to the unequalized image, while the bottom row relates to the equalized image. Here are a few interesting things to note about the graphs in question:

- The equalized histogram actually has the same "shape" as the original histogram, but with the relevant sections spread out. Here, the relevant sections are highlighted in gray.
- Color values which occurred with a high frequency had higher slopes on the CDF. This means that, when mapping from domain to range for the CDF of the unequalized image to generate the equalized image, when using integer color values in the domain, not every color value in the range can be output (see the regions highlighted green in the image). For example, we have to cover the range from values 0 to ~125 in the equalized histogram, with only about 25 discrete points. This is why we see a large number of color values which have frequency 0 in the histogram of the equalized image.
- Having colors which occur with a high frequency have a higher slope, making them take up a larger portion of the range than they do in the domain is also, I feel, the intuitive reason why using the CDF as a lookup table leads to a straight line CDF for the equalized image. If we re-express the pipeline illustration shown on the last page in symbolic notation, we get something like:

$$I(x) \rightarrow F(I(x)) \rightarrow \int F(I(x)) \rightarrow I(X) \circ \int F(I(x))$$
$$\rightarrow F[I(X) \circ \int F(I(x))] \rightarrow \int F[I(X) \circ \int F(I(x))]$$

You'll have to forgive me for the awful typesetting, and all the integrals. I wonder if there's a way for us to convince ourselves that that last term is always a line.