Practical 9 - Run Length

Data compression reduces the size of a file to save *space* when storing it and to save *time* when transmitting it. While Moore's law guides that the number of transistors on a chip doubles every 18-24 months, Parkinson's law tells us that data expands to fill available space. The text, images, sound, video, and so on that we create every day is growing at a far greater rate than storage technology.

Binary Run Length Encoding

Binary Compression

Part 1.

```
command: java BinaryDump 40 < 4runs.bin

Output: 0000000000000001111111000000011111111111

Number of bits: 40
```

Part 2.

```
command: java RunLength - < 4runs.bin | java BinaryDump

Output: 0000111100000111
        0000011100001011

Number of bits: 32
```

compression ratio: compressed bits / original bits = 32/40 = 80%

Part 3.

```
command: java RunLength - < 4runs.bin > 4runsrle.bin
Command: java BinaryDump < 4runsrle.bin

Output:
0000111100000111
0000011100001011

Number of bits: 32
```

ASCII Compression

Part 1.

```
command: java BinaryDump 8 < abra.txt
```

```
Output:
0100000101000010
0101001001000001
0100001101000001
0100010001000001
0100001001010010
0100000100100001

Number of bits: 96
```

Part 2.

```
command: java RunLength - < abra.txt | java BinaryDump 8
```

```
Output:
0000000100000001
0000010100000001
0000000100000001
0000010000000001
0000001000000001
0000000100000001
0000001000000001
0000001000000001
0000010100000001
0000000100000001
0000010000000010
0000000100000001
0000010100000001
0000000100000001
0000001100000001
0000001100000001
0000010100000001
0000000100000001
0000010000000001
0000001000000001
0000000100000001
```

```
0000001000000001
0000001000000001
0000010100000001
0000001000000001
0000010000000001

Number of bits: 416
```

compression ratio: compressed bits / original bits = 416/96 = 433%

Why do you think you got this? What is happening?
We notice that run length compression on the string containing ASCII increases the number of bits. This is because for files that do not contain many runs, run length compression may increase the size.

Bitmap Compression

Run Length encoding is widely used for bitmaps because this input data is more likely to have long runs of repeated data (i.e. pixels).

Part 1.

```
command: java BinaryDump 8 < q32x48.bin

Number of bits: 1536
```

Part 2.

```
command: java RunLength - < q32x48.bin > q32x48rle.bin
command: java BinaryDump 32 < q32x48rle.bin

Number of bits: 1144
```

compression ratio: compressed bits / original bits = 1144/1536 = 74%

Question 3. Perform step 1 and step 2 on the file q64x96.bin

Part 1.

```
command: java BinaryDump 8 < q64x96.bin

Number of bits: 6144
```

Part 2.

```
command: java RunLength - < q64x96.bin > q64x96rle.bin
command: java BinaryDump 64 < q64x96rle.bin

Number of bits: 2296
```

compression ratio: compressed bits / original bits = 2296/6144 = 37%

Question 4.

We notice a significant improvement in the compression ratio between the lower resolution and the higher resolution files. The compression ratios are respectively 74% and 37%. This is due to the fact that by increasing the quality of the file, the number of bits as well as the number of runs have increased.