I graded this myself. The median was 112, and the quartiles were 137 and 89; last year’s corresponding values were 120, 132 and 101. There are lots of variables in play, so I have no explanation of any differences, if indeed there is one other than randomness. The colors in the graph below are for parts 1, 2 and 3, reading up from the bottom.

1. **(50 points, 2 each) Short Answers.** Circle the right answer or write it in the space provided.

(a) Alice wants to digitally sign a document and is wondering whether to use AES or RSA. How would you advise her?

- AES is clearly better
- RSA is clearly better
- both work well
- neither would be good

**RSA.** Public-key crypto can be used for signing; secret key crypto can’t. Covered in the Q/A.

(b) In October, the US Supreme Court finally decided the long-running dispute between Oracle and Google.

-- In whose favor did they decide? **Google**

-- What was the basic issue at issue? (3-4 words is enough!) **Copyrightability of APIs.** I was hoping for this specifically but the question wasn’t as focused as it could have been, so generally anything that said “copyright” was ok.

(c) How old is Cecilia, in base 10?
27. \[ 2 \times 13^1 + 1 \times 13^0. \] Done to death in the Q/A; most people got it.

(d) To compile Fortran, C and C++ programs for a specific kind of computer, say a Mac running macOS, I would need three different compilers. How many different assemblers would I need?

0 1 2 3 4 6

1. The same assembler would be the target of all three compilers.

(e) In late 2020, Apple began selling computers with M1 processors that it designed itself and had fabricated by TSMC in Taiwan. Which of these companies would be most financially affected by Apple’s action?

Amazon Facebook Google Intel Microsoft Netflix

Intel, which makes the processors that previous Macs used.

(f) Add these two binary numbers:

\[
\begin{array}{c}
0110101001.00 \\
1111001010110.11 \\
\hline
10000000000.00000
\end{array}
\]

(g) Suppose that I use SHA-1 to compute a cryptographic hash of a message that I am going to send to a friend. Then I change “bwk” to “BWK” in one place in the message and re-compute the hash. What is the relationship between the first cryptographic hash value and the second?

the same 3 bits are different 3 bytes are different about half the bits differ every bit is different

About half the bits differ. If every bit were different, flipping them back would mean that there was in effect no change. Done at length in class, in the Q/A, and in the book.

(h) Put these names into chronological order of when they made the contribution(s) that caused them to be discussed in COS 109, by writing the numbers 1 through 5 on them.

Tim Berners-Lee Bill Gates Dennis Ritchie Edward Snowden Mark Zuckerberg

Ritchie Gates Berners-Lee Zuckerberg Snowden

(i) Each of the following files is exactly 100 MB long and each contains typical information of the type indicated by the filename extension. Which one of these files would likely be smallest after Lempel-Ziv compression is applied to it?

F.gif  F.jpg  F.mpg  F.mp3  F.png  F.txt  F.zip  no way to tell

F.txt. It’s the only uncompressed format, so it would probably wind up 4x smaller.
(j) What kind or level of programming language is being described in this excerpt from David Auerbach’s 2018 book *Bitwise: A Life in Code*? “Store this number here, retrieve this number from there, add or subtract these two numbers, and branch to different bits of code depending on some condition or other.”

**Assembly language**

(k) The *NY Times* reported on 12/10/14 that the US government had recovered “144,336 bitcoins found on computer hardware belonging to the creator of Silk Road”, a site largely devoted to the sale of illegal drugs. Which one of these specific hardware components would be the most likely place to “find” bitcoins?

<table>
<thead>
<tr>
<th>accumulator</th>
<th>bus</th>
<th>cache</th>
<th>CPU</th>
<th>disk</th>
<th>RAM</th>
<th>ROM</th>
</tr>
</thead>
</table>

*disk.* All the others are transient at best.

(l) What ASCII character occurs most frequently in ordinary English text like a novel or a newspaper story? Hint: as we saw in class, it occurs nearly twice as often as the next most frequent character.

**Space.** Not the letter e. Carefully worded as “character”, not letter.

(m) Headline in early December: “Trump's social media site quietly admits it's based on Mastodon. Mastodon had threatened to sue Trump's 'Truth Social' site for allegedly violating its ________________ license.” What two- or three-word phrase belongs in the blank?

**Open-source.**

(n) Amazon.com and the government of Brazil both want to own the top-level domain .amazon. Which one of these organizations is responsible for deciding who gets the domain name?

<table>
<thead>
<tr>
<th>FCC</th>
<th>FTC</th>
<th>ICANN</th>
<th>ITU</th>
<th>UNESCO</th>
<th>USPTO</th>
<th>WIPO</th>
</tr>
</thead>
</table>

*ICANN.* Most people got this.

(o) Suppose that the Princeton admissions office wants to use machine learning to predict the GPA at graduation from Princeton of high-school students who have applied. Which one of these ML techniques would be the most suitable for predicting the ultimate GPA of current applicants? (This is not to suggest that such techniques would necessarily be useful or desirable.)

<table>
<thead>
<tr>
<th>supervised learning</th>
<th>unsupervised learning</th>
<th>image recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>clustering</td>
<td>recommendation system</td>
<td>sentiment analysis</td>
</tr>
</tbody>
</table>

**Supervised learning.** It’s basically a numeric prediction problem, vaguely analogous to predicting the prices of houses or artwork.

(p) Zoom uses a point-to-point network architecture where each party in a video conference is connected to all the others through Zoom’s servers. How does the number of connections grow in proportion to N, the number of participants?

<table>
<thead>
<tr>
<th>logarithmically</th>
<th>linearly</th>
<th>N log N</th>
<th>quadratically</th>
<th>exponentially</th>
</tr>
</thead>
</table>

**Linear.** The connections go through Zoom’s servers, not from one party to another.

(q) Which one of these entities would I have to deal with if I want to acquire radio frequency spectrum for a new wireless service in the USA? Circle the correct answer(s).

<table>
<thead>
<tr>
<th>AT&amp;T</th>
<th>FCC</th>
<th>FTC</th>
<th>GCHQ</th>
<th>IETF</th>
<th>NIST</th>
<th>RTFM</th>
<th>WIPO</th>
<th>WTF</th>
</tr>
</thead>
</table>

*FCC.* Most people got it.

(r) From *The Dark Hours*, a 2021 detective story by Michael Connelly: “He also uses __________ as a browser. It encrypts his moves and bounces them all over the world. So he’s anonymous.” Passing over whether this is strictly accurate, what belongs in the blank?
**Tor.** Most people got it. (Connelly’s books are not as good as the Bosch TV series derived from them, in my humble but correct opinion.)

(s) If I wanted to collect all the contents of all the disks on the laptops of the 31 students in COS 109 onto a single disk, which one of these is the smallest amount of space that would be sufficient?

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MB</td>
<td>1 GB</td>
<td>1 TB</td>
<td>1 PB</td>
<td>1 EB</td>
<td>1 ZB</td>
<td>1 YB</td>
</tr>
</tbody>
</table>

1 PB. Most of you have 500 GB disks, so in aggregate that’s around 15 TB.

(t) A few years ago, the web page for CS 152 at Harvard said “Almost every computation in the world (at a rough guess, 15,000,000,000,000,000,000,000,000 instructions every second) was written in some programming language.” If the computers doing these computations are all typical laptops like yours, *very roughly* how many computers would that be?

10 billion? \(15 \times 10^{18}\) instructions per second / \(1.5 \times 10^9\) instructions per computer per second.

(u) The ACLU says “Stingrays […] mimic cell phone towers and send out signals to trick cell phones in the area into transmitting their locations and identifying information” rather than sending it to base stations. What kind of attack is this most like?

*man in the middle* ransomware spear-phishing Trojan horse war driving

*man in the middle.*

(v) Modern computers can efficiently process integers of several sizes, usually 1, 2, 4, 8, and sometimes 16 bytes long. Which of these is the least number of bytes that could be used for storing a binary number representing the current population of New Jersey?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
</table>

4. 2 bytes is only 65,000 and 4 bytes is 4 billion; it doesn’t matter what the population of NJ is.

(w) OIT notice, 12/7/21: “The outage that’s currently impacting AWS’s US-EAST-1 region is having a significant impact on the performance of Blackboard and Canvas as well as third-party tools that are integrated with Canvas.” What service is AWS providing to Blackboard, Canvas, et al? A couple of words is enough.

**Cloud computing.** A nice example of one of the drawbacks thereof.

(x) Suppose that Google upgrades the cameras that it uses for Street View from 12 megapixels to 30 megapixels. If Google uses 100 PB to store its existing images for the USA, how much space will it need to store the new images?

250 PB. The image sizes went up by 2-1/2 times.

(y) In the factoring challenge sponsored by RSA Labs, RSA-1024 (not factored yet) is 1024 bits long in binary and 309 digits long in decimal. RSA-2048 (also unfactored) is 2048 bits long. *Approximately* how many digits would it have if written out in decimal?

618, plus or minus about 1.

2. **(20 points) Understanding Programs**

(a) The following Python code is supposed to simulate flipping a fair coin *exactly* 1,000 times. At the end, it should print the number of heads and tails. Sadly, it has several errors and doesn’t work. Fix the errors. You do not need to rewrite it if you clearly indicate the changes you would make.

(This is a question about correct logic; don’t worry about syntax. The expression for computing random numbers is correct: each call of `random.random()` produces a new random floating-point value between 0 and 1. The `print` statement is syntactically correct as well.)
i = 1
heads = 0
print("heads =", heads, "tails =", tails)
while i < 1000:
    r = random.random()  # random number r >= 0, < 1.0
    if r >= 0.5:
        heads = heads + 1
    else:
        tails = 1
    i = i + 1
heads = 0
tails = 0
while i <= 1000:
    r = random.random()
    if r >= 0.5:
        heads = heads + 1
    else:
        tails = tails + 1
    i = i + 1
print("heads =", heads, "tails =", tails)

Basically five things to fix up.

(b) If you want to simulate an unbalanced coin that comes up heads 3/4 of the time and tails 1/4 of the time, what change(s) would you make to the program above to achieve this, after it has been corrected?

\[
\text{if } r \geq 0.25 \ldots \text{ or other more clever variations.}
\]

(c) Suppose that version 2.0 of the Toy machine includes an instruction called \textbf{REM}, which divides its operand into the value in the accumulator and leaves the remainder in the accumulator. For example, if the accumulator contains 17, the instruction \textbf{REM 5} will leave 2 in the accumulator.

What does the following program print when given the sequence of input numbers \textbf{3 1 4 1 5 9 2 6 5 4 0}?

```
TOP   GET
IFZERO BOT if accumulator value is zero, go to instruction BOT
STORE TEMP store accumulator value in location TEMP
REM  2 divide accumulator value by 2, put remainder in accumulator
IFZERO TOP if accumulator is zero, go to instruction TOP
LOAD TEMP load accumulator with value from location TEMP
PRINT print value in accumulator
GOTO TOP go to instruction labeled TOP
BOT   STOP
TEMP 0 when execution begins, this location will contain 0

3 1 1 5 9 5
```

(d) In half a dozen words, what computation is this program performing? Do NOT just repeat the instructions in words.

\textbf{Prints the odd numbers.}

(e) How does the running time grow as a function of or in proportion to \textbf{N}, the number of input numbers?

\textbf{N (linear)}

3. (110 points, 5 each) Miscellaneous
(a) An April 2021 article in the London Review of Books says “a ________________ is a small sequence of letters and numbers that a website generates and deposits in your browser.”

(i) What word belongs in the blank?

cookie

(ii) The article goes on to say “a pixel is a tiny, transparent image, and you can't see it on your screen. […] Why would someone want to show you an invisible image? What does the pixel do?” Very briefly but clearly, answer the author’s question.

When the page is loaded, the browser’s request for the pixel tells the server that the page is being accessed and this can be used for tracking. Lots of people went off on discussions of how pixels make images; that’s clearly not the point when there’s only one pixel and it’s invisible.

(b) Some years ago, a COS109 student told me that the instructions for Lab 3 would not print properly when she used Safari. I could see nothing obviously wrong, so I started removing parts of the file (originally about 6,400 lines) to find smaller versions that still exhibited the problem. Describe a systematic general procedure for efficiently locating such a problem, when you have no clue at all about what the problem is. (It turned out to be a missing </a>.) Be brief! I’m looking for an idea, not an essay.

Use a kind of divide and conquer or binary search: see which half of the document causes the problem, discard the other half, and repeat. A gratifying number of people got the idea.

(c) The number 401b30e3b8b5d629635a5c613c791e has 32 hexadecimal digits.

(i) Which of these could it be? Circle all that are possible.

AES key       Ethernet address        IPv4 address        IPv6 address        MD5 hash        SHA-1 hash

AES, IPv6, MD5. The others are the wrong length.

(ii) Briefly explain why this is not a prime number.

Its binary representation ends in zero so it’s an even number. But it doesn’t end in 14! Hex E is indeed decimal 14, but that’s not relevant here, just the fact that it ends in 0.

(d) On 12/3/14, YouTube described an overflow problem in their programming: “We never thought that a video would be watched in numbers greater than a 32-bit integer (2,147,483,647 views)

(i) If the value 2,147,483,647 is stored as a 32-bit integer, what is its leftmost bit: 0 or 1?

0. This is a variant of a problem set question: the leftmost bit is a sign bit, and is conventionally 0 for positive numbers.

(ii) What is its value in hexadecimal?

7F FF FF FF

(iii) When this number is incremented by 1, what is the resulting value in hexadecimal?

80 00 00 00

(e) Here are some hexadecimal values that sort of spell words; the character 0 is actually a zero and 1 is a one.

BA0BAB     COFFEE     DECODE     EFFACE     FACADE     FOOD1E     OFF1CE

(i) If they are interpreted as ordinary 24-bit integer values, which one has the smallest numeric value?

OFF1CE. It starts with zero; none of the others do.

(ii) Which one has the largest numeric value?

FACADE. FA is larger than F0, and all the others have smaller leading digits.

(iii) If instead they are interpreted at 24-bit RGB colors, which one has the least amount of green?

BA0BAB. 0B is smaller than 0D, and all the others are much larger.

(f) A news story says that there are probably about 100 billion planets in the Milky Way.

(i) If astronomers want to give each planet a unique number, using the smallest possible number of bits, how many bits would that number have?
37. $100 \times 10^9$ is about $2^7 \times 2^{30}$

(ii) How many bytes would such numbers occupy?

5. Many people got both parts of this, which is great.

(g) Princeton logs all your Internet connections, including source IP address, the IP address you visit, your Ethernet address, and the Unix standard time (the number of seconds since 1970) at the beginning of the connection and at the end of it.

(i) If IPv4 addresses are used, how many bytes would be required to store this information for one connection, in the most straightforward and conventional representation?

22. $4 + 4 + 6 + 4 + 4$. Unix times are stored in four bytes (as in a recent problem set) and there are both start and end; some people neglected to include the end time.

(ii) If IPv6 were used instead of IPv4, how many bytes would be required?

46. $16 + 16 + 6 + 4 + 4$.

(h) The first half of the first byte of an IP packet contains the version number of the protocol.

(i) Write out the bit patterns that one might most reasonably expect for IPv4 and IPv6.

IPv4 ___________________     IPv6 ______________________

0100, 0110

(ii) What is the largest version number that this scheme allows for, in decimal?

15

(i) Morse code uses combinations of one to five dots and/or dashes to represent letters, digits, and punctuation marks. For example, E is a single dot (·), A is dot-dash (·-), and Q is dash-dot-dot-dot-dash (---·). Suppose you are designing a new version of a Morse-like code, in which every character will consist of some combination of exactly 6 dots and/or dashes. Describe briefly how you would systematically assign uppercase letters and digits to combinations of 6 dots and dashes. Write down enough of your characters, or explain how you would create them, so clearly that there is no ambiguity about your design.

Write them out in binary, with A = 0, B = 1, etc. If you said binary or ascii, that was good enough. People who wrote out other kinds of patterns mostly got there in the end, but at the cost of a lot of time. And, speaking of close reading of the text, the question does say “uppercase”; a surprising number did lowercase as well.

(j) As data travels across the Internet, it is subjected to a fair amount of processing. For each of the following statements, circle the most appropriate answer.

IP packets have serial numbers to ensure that they are processed in the right order    true    false

IP packets that arrive out of order have to be resent    true    false

a long TCP message is broken into multiple IP packets    true    false

Ethernet packets are reassembled into IP packets at each router along the way    true    false

If an IP packet is damaged in transit, error correction bits will restore it    true    false

false    false    true    true    false

(k) This partial Unix directory listing shows size, modification date and time, and filename for five files. Exactly which pair(s) of files do I have to compare byte by byte to determine whether or not they have identical contents?

<table>
<thead>
<tr>
<th>Size</th>
<th>Date</th>
<th>Time</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>347</td>
<td>Oct 29 16:04</td>
<td>f1.doc</td>
<td></td>
</tr>
<tr>
<td>354</td>
<td>Oct 28 16:05</td>
<td>f1.docx</td>
<td></td>
</tr>
<tr>
<td>354</td>
<td>Apr 22 20:03</td>
<td>f1copy.docx</td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>Sep 20 08:51</td>
<td>f2.txt</td>
<td></td>
</tr>
</tbody>
</table>
all pairs of the 3 files that have the same size. The names can be a lie.

(i) The *NY Times* (12/10/18) says that companies are continuously tracking 200 million US cell phones many times per day per phone. Suppose that an average US phone reports its number and its position to an accuracy of one yard or meter 1,000 times/day. *Very roughly* how many terabytes of tracking information are uploaded by all these phones every day in total? Be precise about your assumptions about how information is represented.

**4 TB? 200 * 10^6 * 1000 * (location+number).** The location and number might be 20 bytes total if encoded a bit; there’s a lot of opportunity for compression if one cared.

(m) Charles Babbage’s mechanical computers used decimal arithmetic. Each digit of a number was represented by a wheel with 10 values around its circumference; thus a 12-digit number would require 12 wheels. Imagine that Babbage had taken an early version of COS109, realized the advantages of binary representation, and wanted to build a prototype binary machine that would handle numeric values up to at least one million (decimal).

(i) If Babbage were to use binary wheels (only two values on each wheel) instead of decimal, how many wheels would he need to handle decimal numbers up to one million?

20. 2^20 is somewhat over a million.

(ii) If he were to use hexadecimal wheels instead (16 digits on each wheel), how many wheels would he need for decimal numbers up to one million?

5 wheels, each of which is in effect 4 bits.

(iii) What hexadecimal value would appear on these wheels when representing the largest possible number?

FFFFF. A number of people carefully converted 1 million into hex F4240 or said just “F”; small penalty for these.

(n) A deep-space communications system reports on the health of a piece of equipment by sending a continuous stream of status reports at 1 bit per second. There are three possible status values: OK, High and Low. 98% of the time, the status is OK, while High and Low each occur only 1% of the time. Give an encoding of the three values into three different bit patterns that will minimize the average number of bits sent over a long period of time. Your encoding does not have to use the same number of bits for each status, but there must be no ambiguity about how to decode a sequence of values as they arrive at the receiver.

0 10 11. Lots of confusion here, with plenty of ambiguous encodings, like 0 01 10 or non-minimal encodings like 00 10 11.

(o) A *Mersenne prime* is a prime number of the form 2^n – 1 where n itself is prime, for example 31 = 2^5 – 1. In December 2018, a new Mersenne prime was discovered, the largest known so far: 2^82589933 – 1. It has 24,862,048 digits in its decimal representation.

(i) If it is written out in binary, how many binary digits does it have?

82589933. How many binary digits does 2^5 – 1 have?

(ii) How many of those binary digits are zero?

None. 2^n – 1 is all 1’s in binary, as noted many times in class.

(p) McKinsey reports (11/13/21) that the value of the autonomous vehicles sector went from $10 billion in 2011 to $320 billion in 2021. Assume (unrealistically) that this improvement was a smooth exponential process.

(i) What was the growth rate of value per month during this time?

3%. By the rule of 72: it’s doubling in 2 years = 24 months.

(ii) If the sector continues this exponential rate of progress, in what year will the value be $10 trillion?

2031. Ten doublings = a factor of 1000 and takes 20 years from 2011.
(q) A Slashdot story last year said “There are only 4 billion floating-point numbers, so you can test them all.” The story didn’t say what kind of test, but suppose it is simply to add each possible pair of numbers and see if each sum is correct.

(i) How many such tests would there be? An expression as an integer power of 2 is all that’s needed.

\[ 2^{32} \times 2^{32} = 2^{64}, \text{ though } 2^{63} \text{ is ok too since } a+b \text{ and } b+a \text{ would probably be equivalent. I also accepted “4 billion } ^2 \”, \text{ though it’s clear that 4 billion is an approximation for } 2^{32}. \]

(ii) Double-precision floating-point numbers are 8 bytes long. How many tests would be needed to do the same tests for all double-precision numbers? Again, a simple expression is all that’s needed.

\[ 2^{128} = 2^{64} \times 2^{64}. \]

(r) Supercomputers with lots of processors are often organized as a “mesh” where each processor is connected to its nearest horizontal and vertical neighbors on a rectangular grid. Suppose that there are \( N \) processors, each processor is an identical rectangular box, and the boxes fill a large room from floor to ceiling.

(i) How many connections to neighbors does a typical processor have?

6. Four sides, top, bottom

(ii) How does the total number of connections grow in proportion to \( N \)?

\( N \).

(iii) If technology improves so that the current length, width and height of each processor can be shrunk by a factor of four, about how many processors would now fit in the room?

\( 64 N. \ 4^3. \)

(s) A \emph{NY Times} headline (1/16/13) says “City police plan to put GPS devices in pill bottles to find drugstore thieves.” Naturally the story is light on technical details but it does say that “when a decoy bottle is lifted from a special base it begins to emit a tracking signal.” Assess whether these statements derived from the article are likely to be correct, or are unlikely to be correct.

The locations of stolen pill bottles could be monitored by GPS satellites \hfill likely unlikely

A GPS device in a decoy bottle could send its location to a GPS satellite \hfill likely unlikely

The device could receive enough power from GPS satellites to send a signal \hfill likely unlikely

A GPS device in a decoy could send its location to cellphone base stations \hfill likely unlikely

Spectrum space for the devices would be allocated by the FDA \hfill likely unlikely

unlikely unlikely unlikely likely unlikely

(t) Suppose, not unrealistically, that \( N \) high-tech companies are involved in a bunch of lawsuits.

(i) If each company sues each other company, how does the number of lawsuits grow in proportion to or as a function of \( N \)?

\( N^2. \ Most \ people \ got \ this. \)

(ii) Companies may also band together in groups of various sizes to sue companies that are not in the group; for instance if \( N \) were 4, we might have A suing B, C and D; A and B suing C and D; A, B and C suing D; and so on. If all possible combinations of companies initiate such suits, how does the number of possible lawsuits grow in proportion to \( N \)?

\( 2^N. \ Same \ for \ this \ part. \)

(u) Random quickies.

Bitcoin can’t be used to pay off ransomware demands because it’s anonymous \hfill true false

Your laptop’s Ethernet address changes as you walk from Friend 008 to the Dinky station \hfill true false
A single parity bit can correct a single-bit error in a single byte  true  false
Bitcoin is an example of a cryptocurrency with a relatively stable dollar value  true  false
Rot26 encryption is easier to decrypt than Rot13 is  true  false

A lossless compression algorithm will make some inputs larger, not smaller  true  false
Most DNS queries will access a root server  true  false
Every DNS query accesses a registrar  true  false
If you use HTTPS to access a web site, your ISP does not know which site it is  true  false
If you use Tor to access a web site, your ISP does not know which site it is  true  false