Problem set 1.

Here are some rough solutions.

General observations:

Watch our for unwarranted precision. The number of times that people use proxes varies wildly, so taking a personal answer and multiplying it by 5,000 (itself an approximation) can't make the answer more precise. There is no precise value for miles of roads, so anything computed from that can't have many significant figures. The price of gas varies every day and is different everywhere in the country, so anything more accurate than say $3 or $3.50/gallon is excessively precise. Car mileages vary a lot (and maybe Google uses electric cars, as suggested by several). And so on.

We did not penalize excess precision, but Anika noted it in some cases, as a warning shot across the bow. Please pay attention to this in subsequent problem sets and exams.

Computationally, you can do precise arithmetic throughout a computation and then throw away most of the digits at the end, but it's usually easier and just as good to throw away most of the digits at the beginning. If you use round numbers you can often do the computation in your head or on paper, without a calculator.

A**rithmetic errors** (factors of 1000) or failure to convert miles to feet or fuel to gallons per mile were not uncommon. Minor penalties for those.

Many people said things like "The US is 3000 miles wide and 1500 miles high, so there are 4,500,000 miles of roads." We noted this but with a warning: this only works because of the assumption of one road every mile. **Watch out for confusion between area and linear dimensions**.

Problem 1: [20 pts, 4 each]

(a) Total number of prox card uses in a day?

We were looking for something like "I use mine 10 times a day, and there are 5,000 undergrads, so 50,000 a day".

There should not be too many significant figures in any answer: there is no such thing as a right answer so you should provide round numbers like 350,000 or 400,000. No penalty this time if you provided too much precision, just a warning. I will be more strict later.

(b) Number of bytes of disk space for a transaction? An example transaction?

Maybe 15-20 bytes for your name, 10-20 for a date and time, 5-20 for the dorm and room? It doesn't much matter, though it should be in the general range of maybe 30-60 bytes. For example,

```
Joe College 09162021 2359 Forbes A123
```

is not quite 40 characters. (Spaces count, and you should include your name.)

(c) An idea for compression?

Any scheme that recognizes repeated data is fine, as is anything that replaces a long piece of information with a shorter one.

For example, you could group all the records that share something, like a building or a date or a name, and then have only one instance of the shared information. Or you might replace a long building name with a short numeric value, or each student name with a netid or student number.

Another way to think of it: Put all the data in a spreadsheet, then sort by some column(s). Any column where the sorting yields significant ranges of identical entries is a good target, since you only need to store the unique value once for the range.

(d) Would all the records fit, and why?

We were looking for a vaguely sensible estimate about the number of students over this time period, multiplied by an estimate of the number of days in a year times the number of years.

My estimate:
250 years * 200 days/year * average 4000 ugrads * 10 uses/day
which is about 2 billion uses. If each takes 50 bytes, that's 100GB, so the data would fit comfortably on a laptop disk even without compression.

(e) Gigabytes for a year of pictures of you?

If you use your prox 10 times a day for 200 days and each picture is 1 MB, that's 2000 MB, or 2GB. That might be too low, but it's good enough.

Problem 2:

(a) Miles of road in the US? (Km is ok too)

If the USA is 3000 miles wide x 1500 miles high, there are 1500 E-W roads that are 3000 miles long and 3000 N-S roads that are 1500 miles long. That's 2 * 1500 *

3000 = 9 million. That's probably too high, but anything in the 1M to 10M range seems ok.

Some people didn't explain how they got from area to length of roads. One way is to say that each one-mile square has four miles of road around the edges, but then you have to divide the total by two to avoid double-counting.

(b) Cost of gas?

If we have 5M miles, gas is $4/gallon, a car gets 20 mpg, that's 20 cents/mile, so $1,00,000. Probably anything in the range $100K to $2M.

(c) Disk storage needed, in TB?

5M miles = 25B feet so there are 25 * 10^9 images.
If an image is 1MB, that's 25 * 10^9 * 10^6 bytes = 25 * 10^15 bytes = 25,000 TB.

(d) Cost of disk storage?

Disk is about $20/TB, so $20 * 25,000 so about $500,000.

Problem 3:

(a) How many GB of secondary storage does your computer have?

128, 256, 512 GB, 1TB are typical.

(b) How many GB of RAM?

4, 8, 16 GB are typical.

(c) Pixels on the screen?

1440 * 900 = 1.3 Mp, or whatever.

(d) Speed in Mbps?

Any number should be fine; there's enormous variability.