Theorem 210. (Euclid) There are infinitely many primes.

Proof. Assume (for contradiction) there is only finitely many primes: $p_1, p_2, \ldots, p_n$.
Consider the number $N = p_1 \cdot p_2 \cdot \ldots \cdot p_n + 1$.
None of the p_i divide N (because division of N by any p_i leaves remainder 1).
Thus any prime dividing N is not on our list. Contradiction. □

A few more famous statements concerning the infinitude of primes.

- **Bertrand's postulate**: for every $n > 1$, the interval $(n, 2n)$ contains at least one prime.
  conjectured by Bertrand in 1845 (he checked up to $n = 3 \cdot 10^6$), proved by Chebyshev in 1852
- **Twin prime conjecture:** there exist infinitely many twin primes, that is, numbers p such that both p and $p+2$ are prime

  Just making sure. $(2, 3)$ is the only pair $(p, p+1)$ with p and $p+1$ both prime. (Why?!)

  Some twin prime pairs. $(3, 5)$, $(5, 7)$, $(11, 13)$, $(17, 19)$, $(29, 31)$, $(41, 43)$, $(59, 61)$, $(71, 73)$, $(101, 103)$, ...

  Largest known one: $\underbrace{3756801695685}_{3 \cdot 5 \cdot 43 \cdot 347 \cdot 16785299} \cdot 2^{666669} \pm 1$ ($200,700$ decimal digits; found 2011)

  Twin prime conjecture. Euclid already conjectured in 300 BCE that there are infinitely many twin primes. Despite much effort, noone has been able to prove that in more than 20 centuries.

  Recent progress. It is now known that there are infinitely many pairs of primes (p_1, p_2) such that the gap between p_1 and p_2 is at most 246 (the break-through in 2013 due to Yitang Zhang had $7 \cdot 10^7$ instead of 246).
- **Prime number theorem**: up to x, there are roughly $x / \ln(x)$ many primes

  Another formulation. The prime number theorem is frequently stated by saying that the "probability" for a (large) number x to be prime is $\frac{1}{\ln(x)}$. Show that this follows from the above!

  Solution. If up to x, there are roughly $x / \ln(x)$ many primes, then the "probability" for x to be prime is about $q(x) = \frac{x}{\ln(x)} - \frac{x-1}{\ln(x-1)}$. We need to show that, for large x, $q(x)$ is about the same as $p(x) = \frac{1}{\ln(x)}$. More precisely, we need $\lim_{x \to \infty} \frac{q(x)}{p(x)} = 1$.

  Can your Calculus II superpowers do that? (I needed two applications of L'Hospital.)

Theorem 211. The gaps between primes can be arbitrarily large.

Proof. Indeed, for any integer $n > 1$,

$$n! + 2, \quad n! + 3, \quad \ldots, \quad n! + n$$

is a string of $n - 1$ composite numbers. Why are these numbers all composite!? □

Comment. Notice how astronomically huge the numbers brought up in the proof are!

Example 212. (Riemann hypothesis) The **Riemann zeta function** $\zeta(s) = \sum_{n \geqslant 1} \frac{1}{n^s}$ converges (for real s) if and only if $s > 1$.

The divergent series $\zeta(1)$ is the harmonic series, and $\zeta(p)$ is often called a p-series in Calculus II.

Comment. Euler achieved worldwide fame by discovering and proving that $\zeta(2) = \frac{\pi^2}{6}$ (and similar formulas for $\zeta(4), \zeta(6), \ldots$).

For complex values of $s \neq 1$, there is a unique way to "analytically continue" this function. It is then "easy" to see that $\zeta(-2) = 0$, $\zeta(-4) = 0$, ... The **Riemann hypothesis** claims that all other zeroes of $\zeta(s)$ lie on the line $s = \frac{1}{2} + a\sqrt{-1}$ ($a \in \mathbb{R}$). A proof of this conjecture (checked for the first 10,000,000,000 zeroes) is worth $1,000,000.

http://www.claymath.org/millenium-problems/riemann-hypothesis

The connection to primes.

$$\zeta(s)=\sum_{n\geqslant 1}\frac{1}{n^s}=\prod_p\left(1+\frac{1}{p^s}+\frac{1}{p^{2s}}+\ldots\right)=\prod_p\sum_{n\geqslant 1}\frac{1}{1-\frac{1}{p^s}}$$

This allows certain statements about the zeta function to be translated to statements about primes.
For instance, the (non-obvious!) fact that $\zeta(s)$ has no zeros for $\operatorname{Re} s=1$ implies the prime number theorem.
http://www-users.math.umn.edu/~garrett/m/v/pnt.pdf

Six of the seven Millenium Prize Problems (including the Riemann Hypothesis), for which the Clay Mathematics Institute has offered 10^6 dollars for the first correct solution, remain open.

https://en.wikipedia.org/wiki/Millennium_Prize_Problems

Comment. Grigori Perelman solved the Poincaré conjecture in 2003 (but refused the prize money in 2010).
https://en.wikipedia.org/wiki/Poincaré_conjecture

Example 213. (P vs NP) P versus NP is another one of the Millennium Prize Problems.

"*If the solution to a problem is easy to check for correctness, is the problem easy to solve?*"
https://en.wikipedia.org/wiki/P_versus_NP_problem
Roughly speaking, consider decision problems which have an answer of yes or no. P is the class of such problems, which can be solved quickly. NP are those problems, for which we can quickly verify that the answer is yes if presented with suitable evidence.
For instance.

- It is unknown whether factoring (in the sense of does N have a factor $\leqslant M$?) belongs to P or not.
- Deciding primality is in P (maybe not so shocking since there are very fast deterministic algorithms for checking primality; not so for factoring)
- The travelling salesman problem is known to be NP-hard, meaning that it is in NP and as "hard" as possible (in the sense that if it actually is in P, then P=NP).

Comment. "Quickly" means that the problem can be solved in time polynomial in the input size.
Take for instance computing $2^n \pmod n$, where n is the input (it has size $\log_2(n)$). This can be done in polynomial time if we use binary exponentiation (whereas the naive approach takes time exponential in $\log_2(n)$).
Comment. This is one of the few prominent mathematical problems which doesn't have a clear consensus. For instance, in a 2012 poll of 151 researchers, about 85% believed P≠NP while about 10% believed P=NP.

12.3 Elliptic curve cryptography

The idea of Diffie–Hellman (used, for instance, in DH key exchange, ElGamal or DSA) can be carried to algebraic structures different from multiplication modulo p.

Recall that the key idea is, starting from individual secrets x, y, to share g^x, g^y modulo p in order to arrive at the joint secret $g^{xy} \pmod p$. That's using multiplication modulo p.

One other such algebraic structure, for which the analog of the discrete logarithm problem is believed to be difficult, is elliptic curves.

https://en.wikipedia.org/wiki/Elliptic_curve_cryptography
Comment. The main reason (apart from, say, diversification) is that this leads to a significant saving in key size and speed. Whereas, in practice, about 2048bit primes are needed for Diffie–Hellman, comparable security using elliptic curves is believed to only require about 256bits.

For a beautiful introduction by Dan Boneh, check out the presentation:
https://www.youtube.com/watch?v=4M8_Oo7lpiA