Stocks, Flows, and All That, or, Levels of Variables and Rates of Change, with some Reminders from Calculus

36-467/667, Data over Space and Time

Fall 2020

Contents

1 How Much vs. How Fast

It's often important to distinguish between variables which say *how much* of something there currently is, and those which say *how quickly* that quantity is changing. In some fields, the "how much" variables are called **stocks** and the "how quickly" are **flows**, while in others they are **levels** and **rates** (or **rates of change**), and in yet others **charges** and **currents**, or **quantities** and **fluxes**.

Regardless of the name, one key to keeping these things straight is to think about the dimensions we use to express them. A "how much" might have dimensions of money, or quantity, or mass, or energy (and so units of dollars, or persons, or kilograms, or joules), but a "how quickly" has the dimensions of its "how much" *per unit time*: it's dollars *per year*, or persons *per week*, or kilograms *per second*.

When time is discrete, so we have times $t, t+1, \ldots t+2, \ldots$, people often use the symbol Δ (capital Greek delta) as a prefix, to indicate rate of change, $\Delta x(t) \equiv x(t) - x(t-1)$. (Note: some people would set $\Delta x(t) = x(t+1) - x(t)$; this isn't a huge difference, but it can lead to annoying off-by-one errors when combining formulas from different sources.) In continuous time, we just use the time derivative, $\frac{dx}{dt}(t)$ or $\dot{x}(t)$ or $x'(t)$. Remember that $\dot{x}(t)$ is, by definition, $\lim_{h \rightarrow 0} \frac{x(t)-x(t-h)}{h}$, so you can see that there's a close relationship between $\dot{x}(t)$ and $\Delta x(t)$.

Rates of change can often be broken down into a sum of contributions from different processes, each of which would have its own impact on the level variable. In these situations, we often speak of each contribution as its own rate, and the over-all sum as the **net rate**. (Sometimes we even group contributions from related processes together and talk about their net rate.) That's a vague, abstract statement, but hopefully the next few examples will clarify.

1.1 Examples of Stocks and Flows / Levels and Rates

1. The volume of water in your bathtub at time t, $V(t)$, is a "how much", a stock or level. (The units are units of volume, gallons or liters or cubic meters.) The net rate at which water is being added to or removed from the tub is the "how fast", flow or rate variable, say dV/dt in continuous time, and it has dimensions of volume per unit time, say gallons per minute or liters per second. This rate of change is the three components:
   a. The rate at which water is flowing in from the tap or spigot;
   b. The rate at which water is flowing out from the drain; and
   c. The rate at which water is splashing over the side of the tub.
2. The number of people in a country at a particular moment in time t, say $n(t)$, is a "how much", a level. This will generally be different from one year to the next, so $n(t) \neq n(t-1)$. The rate of change in population will be $\Delta n(t) = n(t) - n(t-1)$ people per year. Note that this rate of change itself is the sum of a number of different rates of change:
   a. The number of births;
   b. The number of deaths;
   c. The number of people entering the country (immigration);
   d. The number of people leaving the country (emigration). The sum of (c) and (d) would be the net migration rate. — It's worth noting that when people talk about "the birth rate" or "the death rate", they usually mean "per capita" rates, which have been normalized by dividing by the total population, so that if $d(t)$ people die over the course of year t, the death rate would be reported as $d(t)/n(t)$. This would have dimensions of $([people]/[time])/[people])$ or $1/[time]$, but it's usually reported as so many deaths per year per person, or deaths per year per 100,000 people[1].

3. The number of people in a particular location with Covid-19 at a particular time t, say $I(t)$, is a "how much", level. It has units of people. The number of *new cases* per week is, as the phrase suggests, part of the rate of change in $I(t)$. The other contributions to the rate of change include the rate of recovery, the rate of death, and the net rate of migration of infected people.

4. "Wealth" is how much money a person or organization has in total. This has units of currency, dollars or euros or dinars as the case may be. "Income", on the other hand, is how much money the person or organization makes over some period of time, say a week or a year, so it has units of currency per unit time, say dollars per year. Wealth is a stock; income is a flow. The rate of change of wealth is income minus expenditures.

1.2 Example with Population

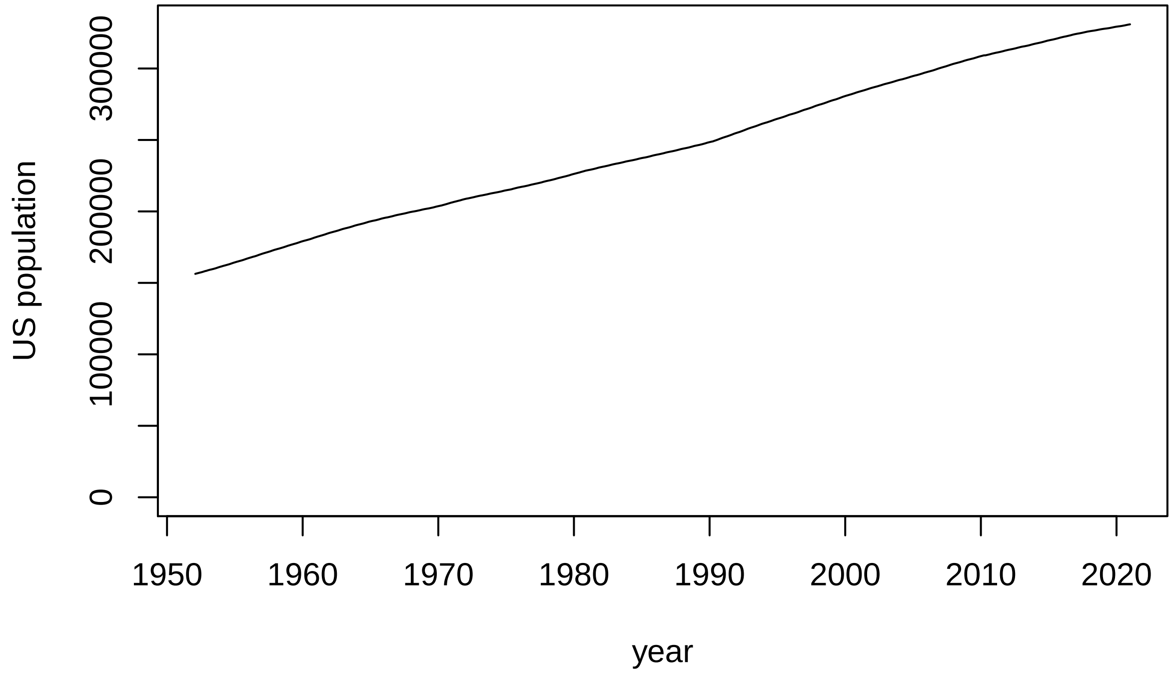
Here is the population of the United States over time[2]:

```r
library(pdfetch)
# Total US population
pop.fred <- pdfetch_FRED("POP")
# This comes as a complicated data type, so break it down to a simple data
# frame, converting dates to a year with a decimal fraction
library(xts) # Functions needed for pdfetch's preferred format
```

[1] You might ask why $d(t)/n(t)$ rather than say $d(t)/n(t+1)$ or even $\frac{d(t)}{(n(t)+n(t+1))/2}$. The short answer is that this is pretty much the probability of a random person in the population at the start of the year dying over the course of the year, which is a good thing to know. The longer answer is that if we're looking at the population of a country or even a city from one year to the next, $n(t)$ and $n(t+1)$ will typically be so close to each other that it really doesn't matter which one we use as the denominator. (That argument breaks down for very small populations, e.g., a single city block. It also breaks down if the time between measurements is long compared to the typical life-span of individuals in the population, e.g., a century for human beings, or a year for fruit flies.) The yet longer answer is that you could use any of these denominators, or others, so long as you were consistent about it. The demographers whose job it is to calculate and use these "vital rates" have names for the different denominators, and elaborate book-keeping schemes for ensuring consistency (clearly explained in, for example, Alho and Spencer (2005)).

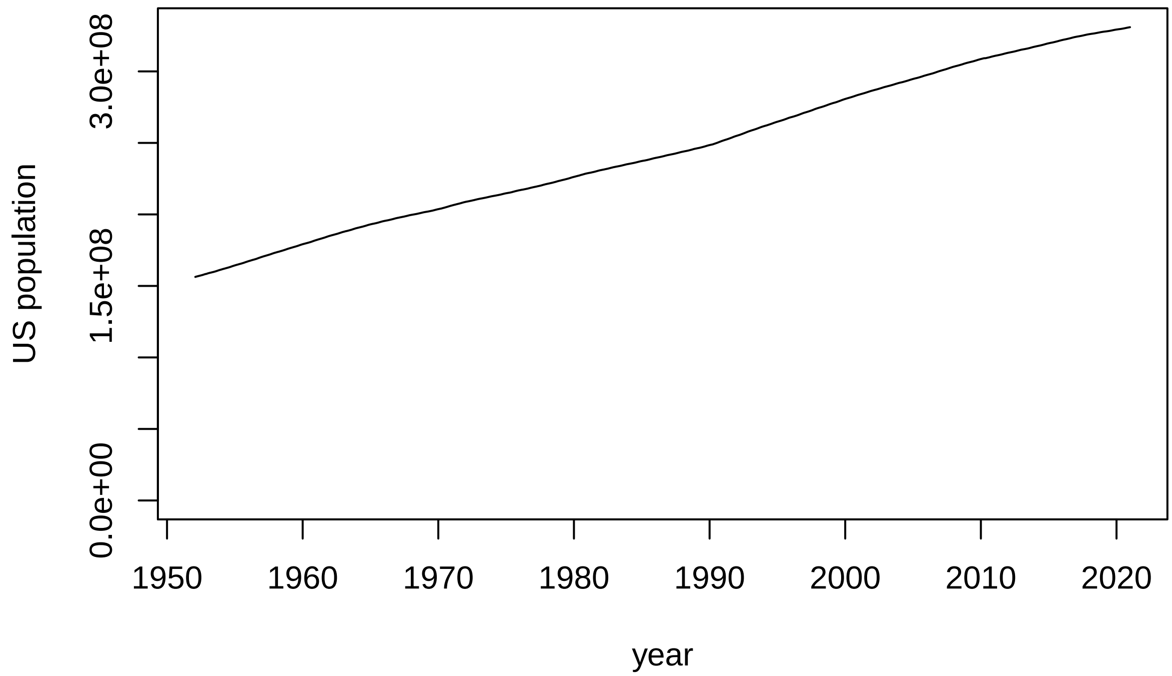
[2] On the mysteriously-helpful FRED data source, see below.

```r
library(lubridate) # Provides useful date-conversion functions
pop <- data.frame(year=decimal_date(index(pop.fred)),
                  y=as.numeric(pop.fred))
plot(pop, type="l",ylab="US population", ylim=c(0, max(pop)))
```



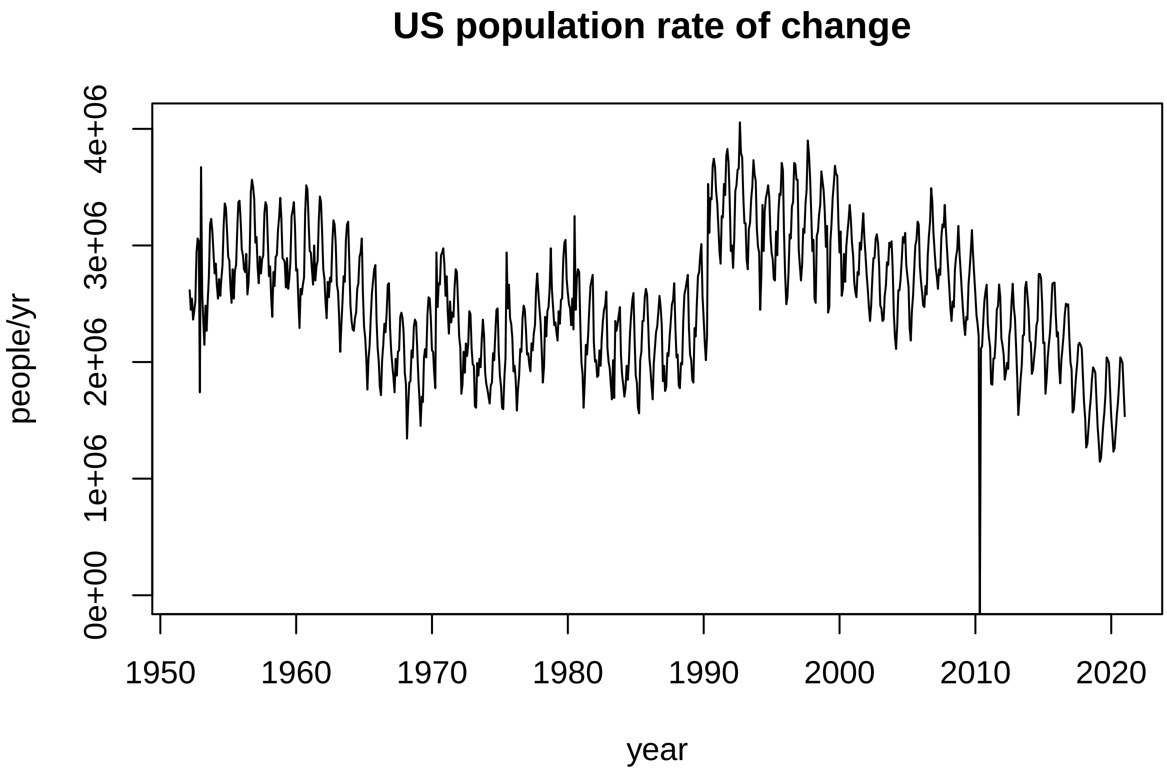
Now at this point you should notice something funny about the plot: the largest number on the vertical axis is only 300,000, which is *much* smaller than the actual population of the US. (It's about the population of the city of Pittsburgh!) The data series is, in fact, in units of *thousands* of people (see [https://fred.stlouisfed.org/series/POP]), not individuals — kilohumans, as it were. We could just live with this, but I prefer to convert the numbers:

```r
pop$y <- 1000*pop$y
plot(pop, type="l",ylab="US population", ylim=c(0, max(pop)))
```



Differencing this gives the change in population from one observation to the next. The time between observations is a month, so we could treat this difference as the rate in people per month. It's a bit easier to grasp though if we convert to people per year, using $\frac{x\text{ people}}{\text{month}}\frac{12\text{ months}}{1\text{ year}}$:

```
plot(pop$year[-1], diff(pop$y)*12,
     xlab="year",
     main="US population rate of change ",
     ylab="people/yr",
     ylim=c(0, 12*max(diff(pop$y))), type="l")
```

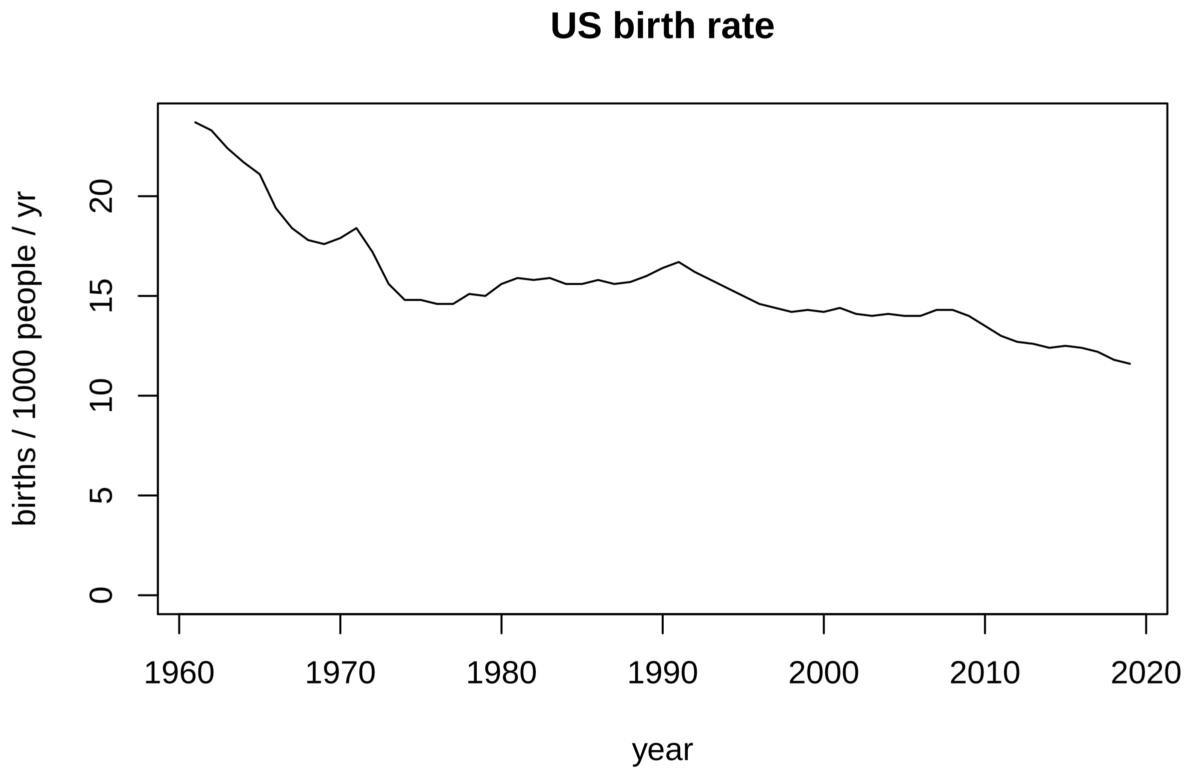


(What, according to the plot, happened around 2010? What do you think *actually* happened then?)

Notice that while the plot of population over time looks like it has a pretty constant slope, that growth looks a lot more erratic on a month-to-month basis. (Though there are clearly some repeating patterns in the rate of change.) Generally speaking, "stock" or "level" variables will have smoother-looking time series than their correspond "flow" or "rate" variables, in just this way.

The same data source gives us access to the birth rate:

```
births.fred <- pdfetch_FRED("SPDYNCBRTINUSA")
births <- data.frame(year=decimal_date(index(births.fred)),
                     y=as.numeric(births.fred))
plot(births, type="l",
     main="US birth rate",
     ylab="births / 1000 people / yr",
     ylim=c(0, max(births$y)))
```



If we had the death rate, we could deduce from this the (net) migration rate. (How?) If, on the other hand, we had the figures for immgiration and emigration, we could deduce the death rate. (Again, how?)

2 Rates from Levels, Levels from Rates

If we know a level variable $x(t)$, we get the corresponding rate by taking differences:

$$\Delta x(t) = x(t) - x(t-1)$$

The appropriate R command, as we've seen in passing, is `diff()` (for "difference").

If we know a rate variable, we can *almost* get the levels back:

$$x(T) = x(0) + \sum_{t=1}^{T} \Delta x(t)$$

The appropriate R command is `cumsum()` (for "cumulative sum"). Notice that we have to know the initial level $x(0)$.

(Of course there's nothing magic about time 0. If we know $x(t_1)$ and want $x(t_2)$ with $t_2 > t_1$, it's $x(t_1) + \sum_{s=t_1+1}^{t_2} \Delta x(s)$. [What would we do if $t_2 < t_1$?])

2.1 Calculus

You may be recalling some formulas from calculus at this point. If we know the function $x(t)$, then the derivative is[3]

$$\dot{x}(t) = \frac{dx}{dt}(t) = \lim_{h \downarrow 0} \frac{x(t) - x(t-h)}{h}$$

Going the other way is integration:

$$x(T) = x(0) + \int_0^T \dot{x}(t) dt$$

Heuristically[4]: $\dot{x}(t)$ is the rate of change in x at time t. So $\dot{x}(t)dt$ is the *amount* of change over the very small time from t to $t + dt$. (Sanity check: $\dot{x}(t)$ has dimensions of stuff per unit time, dt has dimensions of time, so $\dot{x}(t)dt$ has units of stuff.) Adding up all the changes over small intervals of time has to give us the net change, or

$$x(T) - x(0) = \int_0^T \dot{x}(t) dt$$

Because of these relationships, if we start with a time series $x(t)$ and get $Y(t)$ from it by taking cumulative sums, $y(t) \equiv \sum_{s=0}^{t} x(s)$, it's common to talk about $y(t)$ as an *integrated* series.

You may, at this point, recall that in calculus we defined the average value of a function x over the interval from 0 to T as

$$\frac{1}{T} \int_0^T x(t) dt$$

So the *average rate of change* would be

$$\frac{1}{T} \int_0^T \dot{x}(t) dt$$

or

$$\frac{x(T) - x(0)}{T}$$

The latter formula of course also works in discrete time.

3 Relative Change and Growth Rates

Sometimes we are more interested not in the *total* amount by which some quantity changed, but in how big that change was relative to the starting value. That is, instead of caring about

$$x(t) - x(t-1)$$

we might care about

$$\frac{x(t) - x(t-1)}{x(t-1)}$$

[3] If you know enough mathematical analysis to be bothered by my lack of rigor over the next few paragraphs, you also know enough analysis to insert the necessary regularity conditions on the function $x : \mathbb{R} \mapsto \mathbb{R}$ to make it true, and perhaps even to work out the extension to functions which have, say, a finite number of points of discontinuity. (In other words: trust me, I do actually know math.)

[4] In math, a "heuristic" is a way of reasoning which we know isn't quite right, but which often helps us discover true things, which we can then check more rigorously. (It's the same route as "Eureka!", "I have found it!"). Here the heuristic is "when we see expressions like dx or dt in bits of calculus, treat them like arbitrarily small but not quite zero quantities on which the usual rules of algebra apply". This isn't right, "infinitessimals" don't really exist[^nonstandard], but it's a lot easier to follow than a truly rigorous proof of the fundamental theorem of calculus! — The book which introduced the notion of mathematical heuristics, and still one of the best books on the subject, is Polya (1957), which I strongly urge reading. Bressoud (2019) tells the story of how we moved from actually thinking that things like dt had to be infinitessimal numbers, somehow, to modern, rigorous versions of calculus based on limits.

This is sometimes called the **relative** change, and sometimes written $\delta x(t)$ (a lower-case Greek letter delta). Note that the dimension here is (stuff)/(stuff) *per unit time*, or 1/time, whatever the dimensions of x itself might be (mass, people, viruses, money...).

Just as we can recover the new value from the old value and the change,

$$x(t) = x(t-1) + \Delta x(t)$$

we can recover the new value from the old value and the relative change,

$$x(t) = x(t-1)\,(1 + \delta x(t))$$

We've just seen that additive changes, well, add up over time:

$$X(T) = x(0) + \sum_{t=1}^{T} \Delta x(t)$$

Relative changes *multiply*:

$$X(T) = x(0) \prod_{t=1}^{T} (1 + \delta x(t))$$

3.1 Relative Change, Exponential Growth, Exponential Growth Rates

Suppose that each level is some fixed percentage bigger (or smaller) than the one before, so that $\delta x(t) = r$ for all t. Then

$$x(t) = x(0)(1+r)^t$$

Because this involves exponentiating $1+r$, it's called **exponential growth**. People sometimes also use this phrase when $r < 0$, but if we want to distinguish that case, it's called **exponential decay**. r is called the **growth rate**, or, if $r < 0$, we sometimes say $-r$ is the **decay rate**. Thus we might say that "per-capita income grows at 2% per year" or "the amount of radioactive material remaining decays at 5% per year".

This suggests how we should define *average* growth rates:

$$\overline{r} = \left(\frac{x(t)}{x(0)} - 1\right)^{1/t}$$

because then $x(t) = x(0)(1+\overline{r})^t$.

For various reasons[5], it's often more convenient to re-write multiplicative growth as

$$x(t) = x(0)e^{\lambda t}$$

or

$$x(t) = x(0)e^{t/\tau}$$

In the former case, where we write $e^{\lambda t}$, it's easy to convince yourself that

$$\lambda = \ln{(1+r)}$$

This is called the **logarithmic growth rate**, though many people drop the "logarithmic" when they think it's clear from context. (Also, if $|r| \ll 1$, then $\ln{(1+r)} \approx r$, as you can prove using Taylor approximation.) In the latter case, where we write $e^{t/\tau}$, it's easy to convince yourself that

$$\lambda = \frac{1}{\tau}$$

[5]Like if we want to do any kind of calculus.

$|\tau|$ is then called the **characteristic time** of the exponential growth (or decay). It's the time needed for $x(t)$ to change by a factor of e.

Since changing by a factor of e isn't that intuitive, we sometimes convert to using a base of 2 for our exponents,

$$x(t) = x(0)2^{t/\tau_2}$$

$|\tau_2|$ is called the **doubling time** or **halving time** (depending on whether $r > 0$ or $r < 0$); if $r < 0$, it's also called the **half life**. To find τ_2 in terms of r, let's set this equal to our expression in terms of e, simplify, and then take natural log of both sides:

$$\begin{aligned}
x(0)2^{t/\tau_2} &= x(0)e^{t\ln(1+r)} && (1)\\
2^{t/\tau_2} &= e^{t\ln(1+r)} && (2)\\
\frac{t}{\tau_2}\ln 2 &= t\ln(1+r) && (3)\\
\frac{\ln 2}{\tau_2} &= \ln(1+r) && (4)\\
\frac{\ln 2}{\ln(1+r)} &= \tau_2 && (5)\\
\frac{0.6931472\ldots}{\ln(1+r)} &= \tau_2 && (6)\\
\frac{0.6931472\ldots}{r} &\approx \tau_2 && (7)
\end{aligned}$$

where the last approximation works when $|r| \ll 1$. (A rule of thumb you may have encountered is that the doubling time for an investment is 70 divided by the growth rate as a percentage; that's this formula, approximating $\log 2$ as 0.70 rather than $0.6931472\ldots$ for the sake of mental arithmetic.)

Finally, there is a connection between growth rates, and rates of changes of logarithms[6]. Since

$$x(t) = x(t-1)(1 + \delta x(t))$$

we have[7]

$$\log x(t) = \log x(t-1) + \log(1 + \delta x(t))$$

or

$$\log(1 + \delta x(t)) = \log x(t) - \log x(t-1)$$

Now we can *define* the logarithmic growth rate $\rho(t)$ as $\log x(t) - \log x(t-1)$, and we have

$$\log x(t) = \log x(0) + \sum_{t=1}^{t} \rho(t)$$

In other words, if we just take logs, (logarithmic) growth rates work just like ordinary additive rates of change. In particular, we can find the *average* logarithmic growth rate very simply:

$$\overline{\rho} = \frac{\log x(t) - \log x(0)}{t}$$

Sometimes we don't even have to take the logs explicitly, we can just use a logarithmic scale on the vertical axis.

Exercise: The logarithmic growth rate is $\ln(1 + r)$. Does this relationship still hold between the average growth rate and the average logarithmic growth rate?

[6] I'm writing log rather than ln or $\log_k$ because it doesn't matter what base we use for the logarithm, so long as we use the same base for everything.

[7] If $x(t)$ is a variable with units, you should be a little worried about taking its logarithm — what do log kilograms or log dollars mean? But notice that what we end up with is $\log x(t) - \log x(t-1) = \log \frac{x(t)}{x(t-1)}$, and $\frac{x(t)}{x(t-1)}$ is a dimensionless ratio, which will always have the same numerical value no matter what units we use for x (so long as we consistently use the same units over time). In general, if you start with an equation where the units balance, you can safely take the log of both sides (or exponentiate both sides!) and be confident that there's a way of re-arranging the results so that you only take logs of dimensionless ratios.

3.2 Example: Gross Domestic Product Per Capita

The gross domestic product (GDP) of a country in a certain year is the summed price of all the goods and services produced and bought in that country in that year[8]. Calculating GDP has become a central task of official statistical agencies over the last, say, 70 years, since GDP per capita[9] has become one of the ways in which countries keep score with each other. A country's GDP per capita is its mean income per person. Here is the GDP per capita for the United States, as provided by the Federal Reserve Bank of St. Louis[10], as far back as the data go, after adjusting for inflation[11].

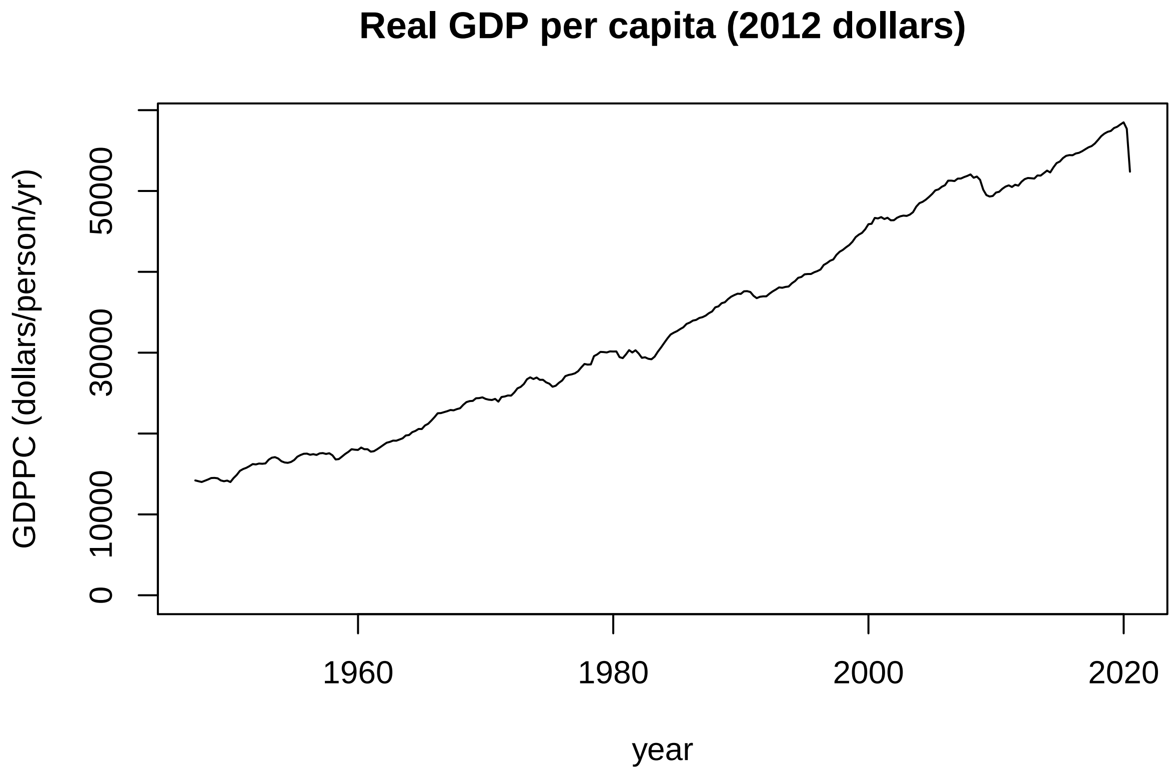
```
library(pdfetch)
# Fetch real (inflation-adjusted) US per-capita gross domestic product
# ("chained 2012 dollars", i.e., no inflation adjustment for 2012)
gdppc.fred <- pdfetch_FRED("A939RX0Q048SBEA")
# This comes as a complicated data type, so break it down to a simple data
# frame, converting dates to a year with a decimal fraction
library(xts) # Functions needed for pdfetch's preferred format
library(lubridate) # Provides useful date-conversion functions
gdppc <- data.frame(year=decimal_date(index(gdppc.fred)),
                    y=as.numeric(gdppc.fred))
plot(gdppc, type="l", ylab="GDPPC (dollars/person/yr)",
     main="Real GDP per capita (2012 dollars)",
     ylim=c(0, max(gdppc)))
```

[8] The difference between gross *domestic* product (GDP) and gross *national* product (GNP) depends on whether we add up all the goods and services produced in a country's territory, no matter who does the work (GDP), or whether we add up all the goods and services produced by a country's citizens, no matter where they are (GNP). They're usually very similar, but can differ if a country's gets of income abroad (because its citizens work abroad, or own assets in other countries, etc.) or conversely has to send a lot of payments to other countries (because lots of its businesses are really shell companies).

[9] "per capita" is a Latin phrase, meaning "for (each) head", that is, "for each person".

[10] The Federal Reserve system, or "Fed", is the U.S.'s central bank, which is (basically) responsible for the money supply for the United States. In most countries, the central bank is, in fact, a single bank, but the Fed is, officially, a federation of a number of regional banks based in cities that were major commercial sites when the Fed was established about 100 years ago, plus the over-all headquarters in Washington, D.C. To do its job of regulating the money supply, the Fed needs to know a *lot* about how the economy is doing, so it's been a pioneer of gathering and analyzing economic data. The regional Fed branches have a lot of independence, and during the early days of the Web, the branch of the Fed in St. Louis, Missouri established a site which provided electronic access to lots of official economic statistics, called FRED ("Federal Reserve Economic Data"), at [https://fred.stlouisfed.org/]. This has continued to today, with a much nicer on-line interface than what we had to deal with when I was your age. (My lawn, please to get off it.) The `pdfetch` library gives an easy way to load these data series in to R, if you know which series you want.

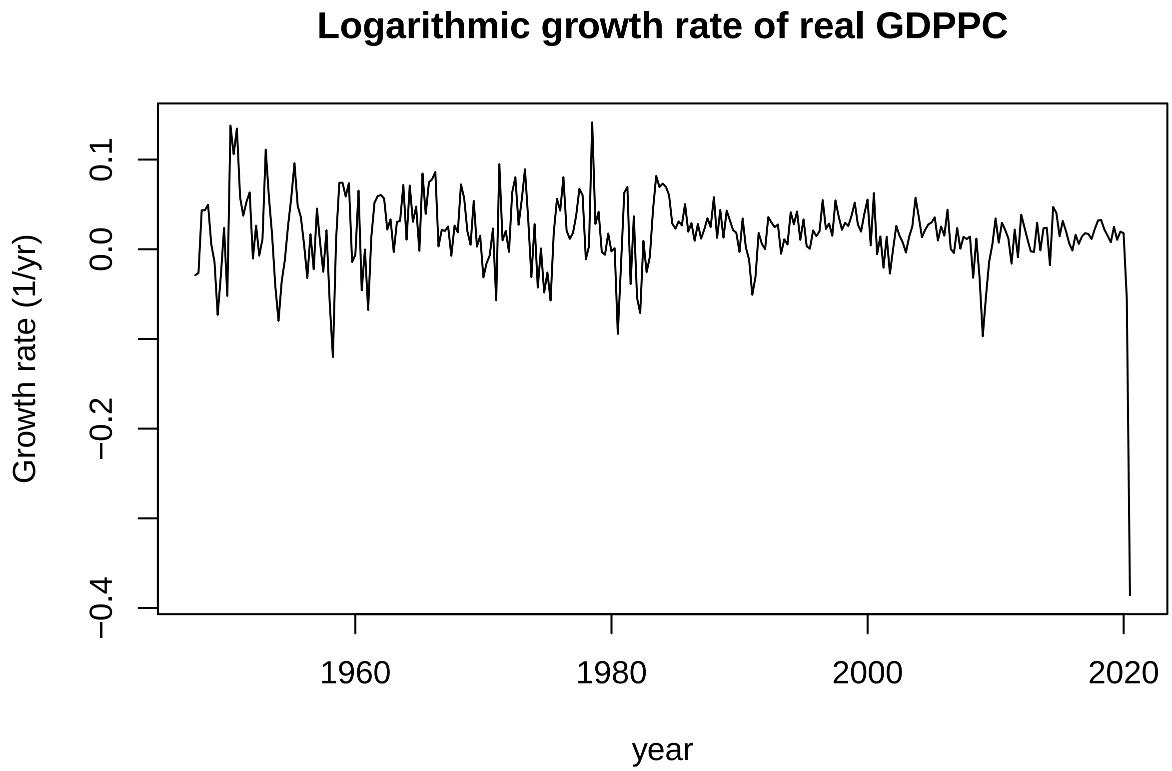
[11] Specifically, this uses the consumer price index (CPI) to go from "nominal" dollars (i.e., the dollars you'd see on a shop's board, or in a catalogue, or on a paycheck) to "real", inflation-adjusted dollars. The base year for the index here is 2012, meaning that's the year when nominal dollars equal real dollars; each nominal dollar in an earlier year (generally) corresponds to more than 1 dollar in 2012, and each nominal dollar in a later year to less than one 2012 dollar. (I could have re-based this to 2019, but it would have been more work than I felt like doing.) — Note the CPI is not the only way to adjust for inflation. There are actually separate CPIs for urban and rural consumers (the over-all CPI averages these), a very different produced price index, etc. There is also a lot of controversy over whether the CPI adequately handles technological changes (how much would it have cost in 1970 to get services equivalent to a low-end 2020 smart phone?) and social changes (college costs a lot more than in 1970, which is easy to handle, but it's also seen as more important, which is harder).



We can look at the logarithmic growth rate by taking difference between successive logarithms of this series[12]:

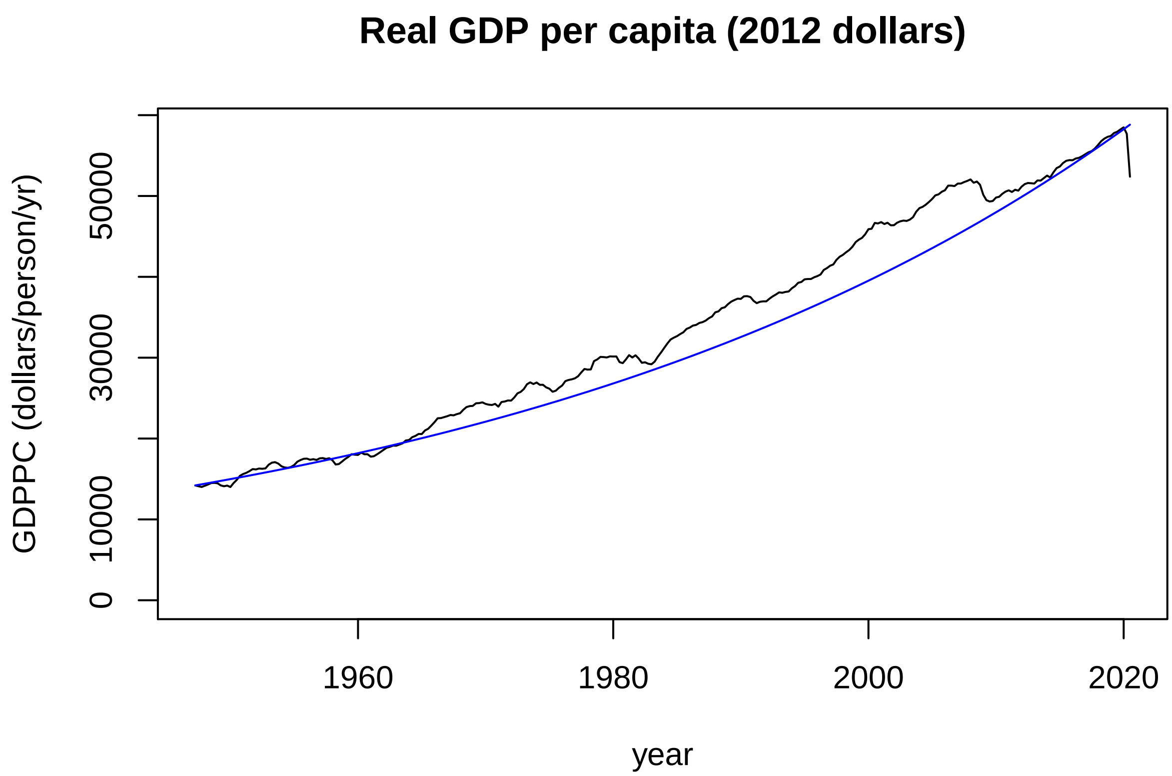
```
plot(gdppc$year[-1], diff(log(gdppc$y))*4, type="l", xlab="year",
  main="Logarithmic growth rate of real GDPPC",
  ylab="Growth rate (1/yr)")
```

[12]Why do I have to omit the first year in the plot above? Why do I multiply by 4?



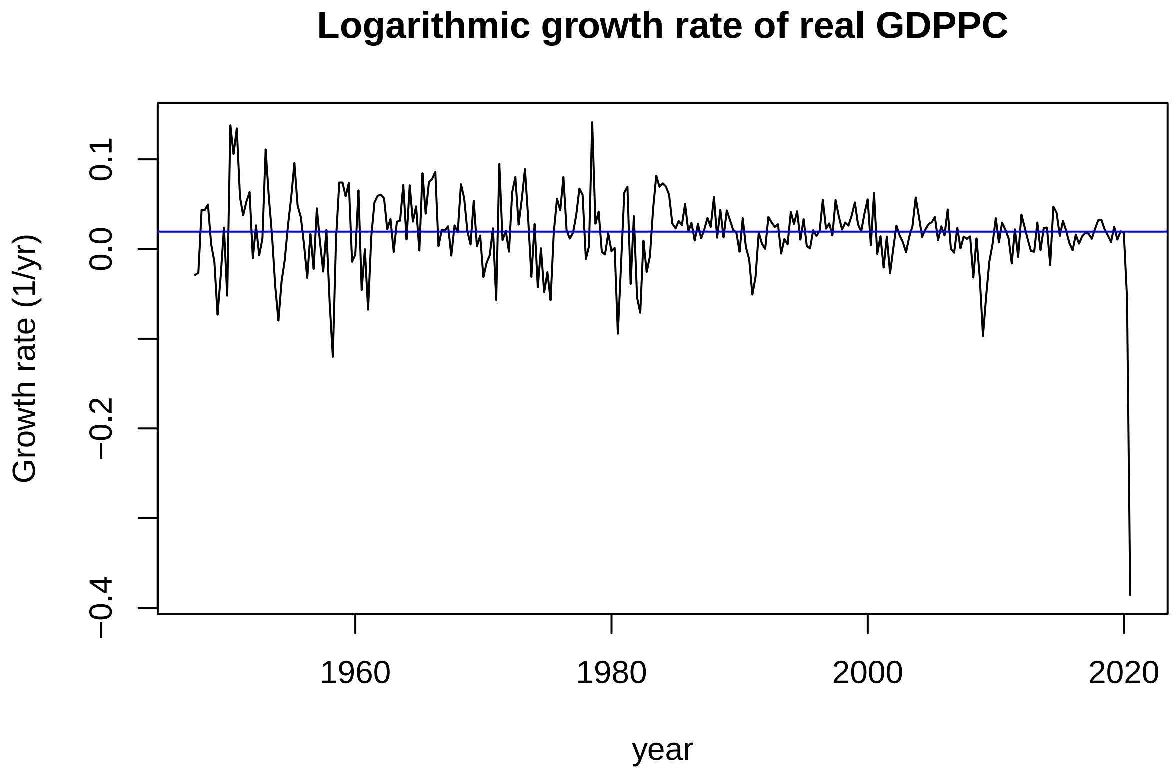
If we omit the very last value from the GDPPC series (why?), we can find an average growth rate, and see that growth was (kind of) steady around that rate:

```
T <- max(gdppc$year) - min(gdppc$year)
avg.log.growth.rate <- (log(gdppc$y)[length(gdppc$y)-1] - log(gdppc$y)[1])/(T-1)
plot(gdppc, type="l", ylab="GDPPC (dollars/person/yr)",
     main="Real GDP per capita (2012 dollars)",
     ylim=c(0, max(gdppc)))
curve(gdppc$y[1] * exp((x-gdppc$year[1])*avg.log.growth.rate),
      add=TRUE, col="blue")
```



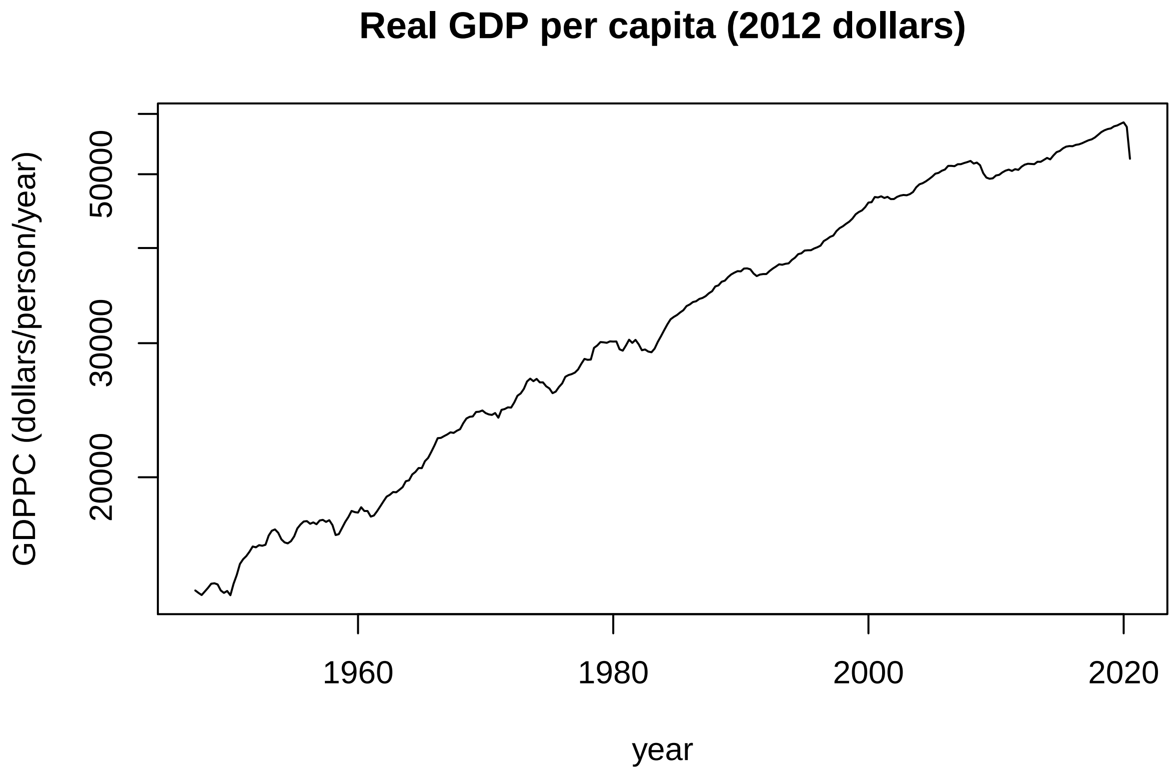
Of course, the actual growth rates center nicely around their average:

```
plot(gdppc$year[-1], diff(log(gdppc$y))*4, type="l", xlab="year",
  main="Logarithmic growth rate of real GDPPC",
  ylab="Growth rate (1/yr)")
abline(h=avg.log.growth.rate, col="blue")
```



Another way to get that average log growth rate would be to just use a logarithmic scale on the vertical axis (without taking any differences):

```
plot(gdppc, type="l", ylab="GDPPC (dollars/person/year)",
     main="Real GDP per capita (2012 dollars)",
     log="y")
```



References

Alho, Juha, and Bruce D. Spencer. 2005. *Statistical Demography and Forecasting.* Berlin: Springer.

Bressoud, David M. 2019. *Calculus Reordered: A History of the Big Ideas.* Princeton, New Jersey: Princeton University Press.

Polya, G. 1957. *How to Solve It: A New Aspect of Mathematical Method.* Second. Princeton, New Jersey: Princeton University Press.