**Motivation**

A couple of weeks back, [Branko Milanovic asked on Twitter](https://twitter.com/ellis2013nz/status/1064276290146955264) :

“Does anyone know a link to a calculation on how many people were born … in the entire 20th century?”

Somewhat surprisingly, no-one did. However, there was a [calculation by the Population Research Bureau](https://www.prb.org/howmanypeoplehaveeverlivedonearth/) that about 108 billion people had walked the earth since 50,000 years ago. I gave myself the detective job of tracking done the data to make the estimates for just the twentieth century.

The necessary data is crude birth rates and total population numbers. Back to about 1950 there is excellent data available from the UN, but before that it is surprisingly hard to find from official sources.

**Kremer’s long range global population estimates**

The first useful thing I found was an interesting article in *The Quarterly Jorunal of Economics* by Michael Kremer on [Population Growth and Technological Change: One Million B.C. to 1990](http://faculty.econ.ucdavis.edu/faculty/gclark/210a/readings/kremer1993.pdf)

That got me this dramatic-looking set of numbers:

The vertical axes in those charts is on a logarithmic scale, so the growth rates in population there are truly astonishing. Humanity has exploded on this planet in a very short period of time, in the scale of things.

Here are just the growth rates:

Some interesting things here include:

* the dip into negative territory in the thirteenth century, with Mongol wars and the Black Death;
* another bad time in the early seventeenth century with the 30 years war destroying Germany, and the collapse of the Ming dynasty in China;
* a dip around 1850 which I think is probably associated with the Taiping Rebellion in China. Despite its massive scale we saw a decline in growth but nowhere near down to negative territory;
* a final dip associated with World War I and subsequent influenza pandemic (years between 1900 and 1920);
* maximum growth rate in 1960 with rapid decline since.

There would be better sources now of the recent years, showing the steep decline continuing. For the record, the world population on current trends in declining growth rates will reach a maximum level of [around 11 billion in 2100](https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html).

I typed out Kremer’s numbers rather than muck around trying to read them electronically. Note that there seems to be an error in the third growth rate in his Table I on page 683; while all his other growth rates exactly match my calculations based on his population levels, for that particular number he has 0.000031 and I get 0.00012. I don’t think it matters much for the substance of his interesting arguments in that paper.

Here’s the R code to create those charts with Kremer’s numbers:

library(tidyverse)

library(viridis)

library(gridExtra)

library(ggrepel)

library(readxl)

library(testthat)

set.seed(123)

#-----------------total population size---------------

kremer <- data\_frame(

year = c(-1000000, -300000, -25000,

-10000, -(5:1)\*1000, -500, -200,

1, 200, 400, 600, 800,

10:16 \* 100,

1650, 1700, 1750, 1800, 1850,

1875, 1900,

192:199 \* 10),

pop = c(0.125,1,3.34,4,5,7,14,27,50,100,150,170,190,190,200,220,

265,320,360,360,350,425,545,545,610,720,900,1200,1325,

1625,1813,1987,2213,2516,3019,3693,4450,5333) \* 1000000

) %>%

mutate(growth = lead((pop / lag(pop)) ^ (1 / (year - lag(year))) - 1))

# note there's an error in Kremer's 3rd growth rate: 0.000031 should be 0.000012

# Global populations

p1 <- kremer %>%

ggplot(aes(x = year, y = pop)) +

geom\_line() +

geom\_point() +

scale\_y\_log10(label = comma, limits = c(1e5, 5e9)) +

scale\_x\_continuous("Year") +

labs(caption = " ")

p2 <- p1 %+% filter(kremer, year > -2000)

grid.arrange(p1 +

labs(y = "Global human population (logarithmic scale)") +

ggtitle("World population",

"1 million BC to 1999"),

p2 +

labs(y = "") +

ggtitle("",

"1,000 BC to 1999") +

labs(caption = "Source: Kremer, 1993, 'Population Growth and Technological Change: One Million B.C. to 1990"),

ncol = 2)

# Global growth rates

kremer %>%

filter(year > -2000) %>%

ggplot(aes(x = year, y = growth)) +

geom\_path() +

geom\_point() +

geom\_text\_repel(aes(label = year), colour = "steelblue") +

scale\_y\_continuous("Annual growth rate", label = percent) +

labs(caption = "Source: Kremer, 1993",

x = "Year") +

ggtitle("World population growth rates",

"1000 BC to present")

**Gapminder’s country-level birth rate and population estimates**

That was all very interesting and gave me at least some benchmark population values covering the whole twentieth century (not just the post-WWII period covered in the official sources), but I also need crude birth rates – how many people born per 1,000 people living. This is the only practical way of getting estimates of the number of births; just growth rates alone won’t do it because the same growth rate could mean quite different birth rates, depending on death rates.

Eventually I realised that [Gapminder](https://www.gapminder.org/) publish estimates of many variables, including basic demographic data, back to 1800 at the country level. Gapminder Foundation is the Swedish NGO founded by the recently deceased and much missed Hans Rosling and friends and family, promoting increased understanding of development issues in a historical context. It’s possible to combine these to get global estimates by summing up population numbers, and creating population-weighted averages for birth rates.

This process gives me this plausible-looking set of estimates of the world’s crude birth rate over the last 200 years or so:

These numbers are close enough (for my purposes) to the [UN’s figures](https://en.wikipedia.org/wiki/Birth_rate) for the overlapping period from 1950 onwards.

I also quite like this representation of the population and the birth rate together as a connected scatter plot:

Obviously, once we have crude birth rates and population numbers, we just need to multiply them together to get an estimated number of births per year:

What’s with that jump in the 1980s? Well, the decline in crude birth rate stalled at around 27 or 28 for 15 years or so, and with the massive increase in population coming from rising living standards post-WWII that was enough for a major increase in number of babies being born. By the 1990s, crude birth rates resumed a precipitous decline. There also might be a story here about demographic collapse in post-Soviet Union countries; or there might be a quirk in the data arising from how I aggregated up the country level data.

Here’s the code to get the data from Gapminder and draw those charts:

#---------------birth rate---------------

# harder

# See https://ourworldindata.org/fertility-rate

# nothing before 1950

# The gapminder R package only has data for every 5 years, from 1952

# but the gapminder website has the full data

# crude birth rate per country per year

if(!file.exists("cbr.xlsx")){

download.file("https://docs.google.com/spreadsheet/pub?key=tUSeGJOQhafugwUvHvY-wLA&output=xlsx",

destfile = "cbr.xlsx", mode = "wb")

}

cbr\_orig <- read\_excel("cbr.xlsx")

names(cbr\_orig)[1] <- "country"

cbr <- cbr\_orig %>%

gather(year, birth\_rate, -country) %>%

mutate(year = as.integer(year))

# population per country per year

if(!file.exists("pop.xlsx")){

download.file("https://docs.google.com/spreadsheet/pub?key=phAwcNAVuyj0XOoBL\_n5tAQ&output=xlsx",

destfile = "pop.xlsx", mode = "wb")

}

pop\_orig <- read\_excel("pop.xlsx")

names(pop\_orig)[1] <- "country"

pop <- pop\_orig %>%

gather(year, total\_population, -country) %>%

mutate(year = as.integer(year))

combined <- cbr %>%

full\_join(pop, by = c("year", "country")) %>%

arrange(country, year) %>%

mutate(births = total\_population \* birth\_rate / 1000) %>%

filter(!is.na(total\_population))

ave\_br <- combined %>%

filter(!is.na(birth\_rate)) %>%

group\_by(year) %>%

summarise(birth\_rate = sum(birth\_rate \* total\_population) / sum(total\_population),

total\_population = sum(total\_population)) %>%

mutate(year\_lab = ifelse(year %% 50 == 0, year, ""),

people\_born = birth\_rate \* total\_population / 1000)

set.seed(123)

ave\_br %>%

ggplot(aes(x = year, y = birth\_rate)) +

geom\_line() +

geom\_point() +

geom\_text\_repel(aes(label = year\_lab), colour = "steelblue") +

labs(caption = "Source: Estimated from Gapminder country level data for crude birth rate and population",

x = "Year",

y = "Estimated global crude birth rate per 1,000 population") +

ggtitle("Estimated global crude birth rate",

"1800 to present")

ave\_br %>%

ggplot(aes(x = total\_population / 1e6, y = birth\_rate, label = year\_lab)) +

geom\_path() +

geom\_text\_repel(colour = "steelblue") +

scale\_x\_continuous("Estimated global population", label = comma\_format(suffix = "m")) +

labs(y = "Estimated global crude birth rate per 1,000 population",

caption = "Source: Estimated from Gapminder country level data for crude birth rate and population") +

ggtitle("Estimated global crude birth rate",

"1800 to present")

ave\_br %>%

ggplot(aes(x = year, y = people\_born / 1e6)) +

geom\_line() +

scale\_y\_continuous("People born per year", label = comma\_format(suffix = "m")) +

labs(caption = "Source: Estimated from Gapminder country level data for crude birth rate and population",

x = "Year") +

ggtitle("Estimated global births",

"1800 to present")

**Checking against a more definitive source for more recent years**

I was a bit worried about that bump in the 1980s, so I thought I should have a look at a more definitive data source for recent decades rather than relying on my population-weighted average of Gapminder’s country level data. I grabbed the World crude birth rate from the World Development Indicators for 1960 onwards:

Unsurprisingly it’s a bit flaky in the early years, but I decided the two were close enough that I could stick to using my Gapminder estimates. Here’s how I got that World Bank data:

#------------------better source for more recent years----------

library(WDI)

wdi\_cbr <- WDI(country = "1W", indicator = "SP.DYN.CBRT.IN", start = 1950, end = 2020)

CairoSVG("..http://freerangestats.info/img/0141-compare.svg", 7, 6)

wdi\_cbr %>%

select(year, SP.DYN.CBRT.IN) %>%

rename(wdi = SP.DYN.CBRT.IN) %>%

left\_join(ave\_br, by = "year") %>%

mutate(year\_lab = ifelse(year %% 5 == 0, year, "")) %>%

ggplot(aes(x = birth\_rate, y = wdi, label = year\_lab)) +

geom\_abline(slope = 1, intercept = 0, colour = "orange") +

geom\_path() +

geom\_text\_repel(colour = "steelblue") +

labs(x = "Population-weighted average of Gapminder country data",

y = "World Bank's World Development Indicators") +

coord\_equal() +

ggtitle("Comparing two sources on global birth rates")

**Cumulative births**

Finally, the answer to the question, which turns out to be about 9.75 billion:

The Gapminder data doesn’t have values for every year, but it’s straightforward to interpolate them and get the estimated number of births. That was the final bit of calculating to do.

I saved the [birth numbers from 1800 onwards as a CSV](http://freerangestats.info/data/cumulative-births-1800-onwards.csv) in case anyone is interested in them.

Here’s the code for the final step:

#-----------cumulative births-------------

twentieth\_c <- data\_frame(

year = 1901:2000,

births = approx(ave\_br$year, ave\_br$people\_born, xout = 1901:2000)$y

) %>%

mutate(cum\_births = cumsum(births))

twentieth\_c %>%

ggplot(aes(x = year, y = cum\_births / 1e6)) +

geom\_line() +

scale\_y\_continuous("Cumulative births in the twentieth century", label = comma\_format(suffix = "m")) +

labs(x = "Year",

caption = "Source: Estimates based on Gapminder country level data for crude birth rate and population") +

ggtitle("How many people born in the twentieth century?",

paste("An estimated", format(round(sum(twentieth\_c$births / 1e6)), big.mark = ","), "million"))

# Write a version from 1800 onwards in case people want it:

published\_data <- data\_frame(year = min(ave\_br$year):max(ave\_br$year)) %>%

mutate( births = round(approx(ave\_br$year, ave\_br$people\_born, xout = year)$y)) %>%

mutate(cum\_births = cumsum(births))