Population Accounts

Lars Pedersen, Statistics Greenland, larp@stat.gl

Summary (spoiler alert):

With PxWeb, Nordic NSIs have a world class dissemination tool.

This paper asks: Can an NSI profit from disseminating data?

Answer: Yes, by exposing subject specific, relevant detailed data, support theoretical frameworks, support self-documenting calculations.

Case: Populations Accounts, Greenland

Content

Introduction

Being last on the 'Disseminate' agenda allows me to stand on shoulders of Jan Bruusgaard, a long time Norwegian colleague, with in-depth knowledge of disseminating statistics. He has walked us through the many paths to Nordic free and open data. Since 2013 (improved version in 2019), statistics from the Nordics are all accessible with the same api (except from Denmark); we don't have to invent our own set of tools and at the same time it is so much more interesting for big players with a single, common api. The smaller the NSI, the more benefits we gain, but in a global context, even put together, the Nordics are all small.

Takeaway 1: NSIs in the Nordics are all small ! Some are smaller than others

30 years ago, at the 19th Statistical Meeting in Reykjavik, I presented my study and classification on immigrants and descendants in Denmark – me in the middle –

and if you look closely to the screen of the personal computer from the same meeting, you'll see an early version of PC-Axis in DOS presenting the common Nordic CD 'Statistics Across Borders/Statistics without Frontiers' - and the rest is history, stepstones toward PxWeb of today.
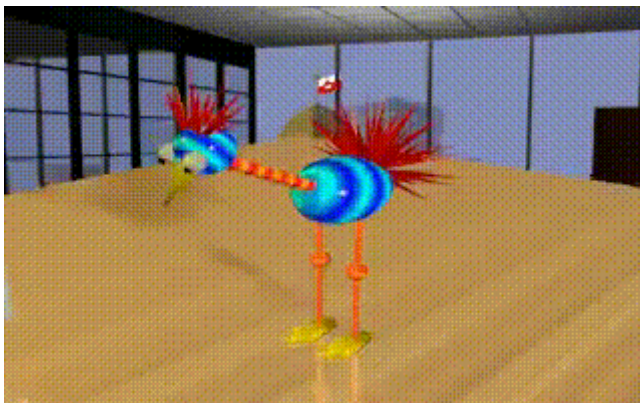
Explorer? Tourist? Farmer?

Curious?

Click here to watch the full 1995 movie

PxWeb in the Nordics

All Nordic Statistical Institutes use the same underlaying Statbank-system and benefits from PxWeb (except Denmark), being a solid and popular generic statistical service adopted also by other NSIs around the globe. Each Nordic NSI has invested heavily in specialized tailormade back-office support for PxWeb to be a vital integrated component in all production flows. It is easy to forget, that what is seen, on internet, only is the tip of the iceberg. Preparing content for our service is what an NSI does. Years ago, NSIs had a rule of thump saying that 60 percent of production time was spent on cleaning raw data, pretty much in line with many survey results.

Over the past 30 years, more than 20 NSIs have co-operated and invested intellectual property into developing the present system, although only few have actually contributed, with code. Like Statistics Greenland, the majority of NSIs have been users of PxWeb offering feed-back and ideas to further development.

Out-of-the-box PxWeb offers a standard user experience, difficult to some, but usage statistics reveal that it is accessible to many. Nevertheless, the Statbank today is integrated as data storage to many alternate presentations, not only on NSI's websites. Either as simple links, charts or tailormade richer presentations from a subset of tables highlighting specific issues. These serve as important showcases for others to replicate. A Greenlandic example is www.stat.gl/demo with code on GitHub, a free and open-source code management system.

Indirectly, user experience is improved by allowing business intelligence software to tap into our Nordic data sources and explore business opportunities of their own. In 2013 PxWeb opened for api access (extended functionality in 2019) that allows developers to reuse and include statistical information in solutions. Look to statistica.com, knoema.com or tableau.com – all US based business intelligence companies, making profits on Nordic data.

Non-profit disseminators like https://www.nordicstatistics.org/ (Nordic co-operation since 1962, Norden.org), today outsourced to Statisticon.se to rationalize data-collection on comparable statistics on the Nordics, and again disseminate with PxWeb. R is used for this process, but collection methods and data wrangling are kept as business secrets.

This is contrary to universities where research traditions aim for transparency and reproducibility. With R as free and open-source software, enthusiasts can build merits creating packages, like the pxweb package. The general available package bridge tables from national statistical institutes directly to statistical analysis, facilitated by the function: pxweb_interactive.

Statistics Greenland has created an r-package, published on GitHub, to ease access to Statbank Greenland and generate standard presentations.

```{r install english statgl}
if (!require("tidyverse")) install("tidyverse")
library(tidyverse)
if (!require("devtools")) install("devtools")
library(devtools)
if (!require("statgl")) install_github("StatisticsGreenland/statgl")
library(statgl)
language <- "en"
```

This is used presenting the Greenlandic SDG indicators on https://stat.gl/en/sdg. A code button reveals all code needed to get hold on data and to replicate presentations.

The saved queries featured in PxWeb are used when linking to data from a presentation. The link takes a reader to the relevant table, showing what data was selected and what else can be selected.

Graph 1. Population size. January 2017 - 2022Q2

Source: Statistics Greenland – See table in StatBank https://bank.stat.gl/BEESTK1
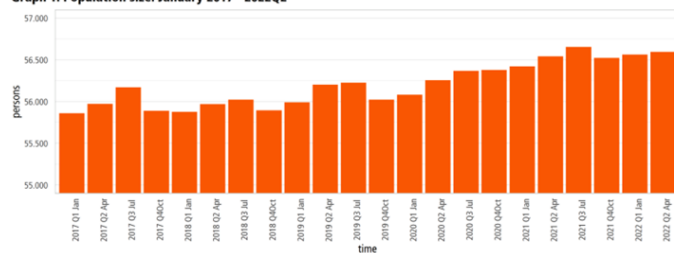
Table 1. The population in municipalities

	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4	2022Q1	2022Q2
Hele landet	56.367	56.379	56.421	56.541	56.653	56.523	56.562	56.595
Kommune Kujalleq	6.458	6.478	6.454	6.409	6.363	6.320	6.292	6.273
Kommuneqarfik Sermersooq	23.211	23.354	23.462	23.632	23.674	23.864	23.861	23.964
Qeqqata Kommunia	9.376	9.305	9.305	9.291	9.355	9.256	9.291	9.275
Kommune Qeqertalik	6.343	6.350	6.312	6.274	6.278	6.176	6.180	6.136
Avannaata Kommunia	10.907	10.818	10.820	10.830	10.914	10.838	10.908	10.887

Source: Statistics Greenland – See table in StatBank https://bank.stat.gl/BEESTK3

With the exception of information on few deaths, information on births and deaths are reported very quickly. Due to a lower number of deaths, the total birth surplus (births minus deaths) was 67 persons in the first 3 months of the year.

The purpose of Statistics Greenland is not to make a profit, but to generate as much prioritized knowledge as possible on our society within a tight budget and total staff of 14.

This paper shares ideas to how a NSI can improve the value of a Statbank service, both for external users but also for a very important user-group, namely ourselves. Out-of-the-box PxWeb is base for the ideas. Nobody lives forever, we do not have to wait for anything to be developed.

Takeaway 2: Nobody lives forever

Not to say, that improvements cannot be made to PxWeb, open-source since august 2019 and free of charge to NSI's. In the following text, Population statistics serve as case.

Population statistics

Population statistics are all about estimating the size of a country's population, its sex and age distribution and development over time. Not only today and yesterday, but also implications for tomorrow and the many days to come.

Population statistics have for at least two millenniums been of vital importance to any administration, whether being a Roman Empire or a contemporary community.

At country level, changes in Population size, is described by the vital statistics: Births, Deaths, Immigrations and Emigrations

Per definition the central equation for a country is

$$P_t + B_{(t,t+1)} - D_{(t,t+1)} + I_{(t,t+1)} - E_{(t,t+1)} = P_{t+1}$$

Only in theory and in a world of perfect statistics will this equity be true. A Population Account brings together statistics on each component and to balance, the account must include a component to hold corrections, in some sources also called 'statistical adjustments'. At subnational level, components for internal migration is added in the Greenlandic regional population account. For the equity above, focus is on national level, to keep it as simple as possible in this paper.

To be really useful and comparable all statistics need to be available detailed by sex and 1-year age groups as all demographic measures show global differences in these dimensions. Fertility, mortality and migrations all vary with age and sex.

In addition, there might be other important country-specific descriptive dimensions to be defined. Over the past 70 years the Nordics have become less homogeneous with regards to the populations' country of birth, as groups of immigrants & descendants have taken permanent residence in the Nordics with various cultural backgrounds. In Greenlandic context 'Place of Birth' is the traditional important demographic characteristic and is integrated into many statistics.

Population statistics supplies base line to all socioeconomic studies. It is pivotal to get it right.

Legacy

The idea of a detailed Population Account leans to a Danish role model. From 1964 to 2004 a detailed Population Account on the Danish Population (sex and 1-year age) was prioritized, to be published annually on paper by Statistics Denmark in "Vital Statistics/(Befolkningens bevægelser)" and as such gave the full and detailed base data for all key figures on survival and other demographic measures. Unfortunately, this table was discontinued in 2005, and so far, has not been continued in Statbank Denmark. Not claiming causality, but the interest in demography in Denmark has been declining, ever since.

Danmark. 01.01.2004 - 26.06.2022.

Google Trends
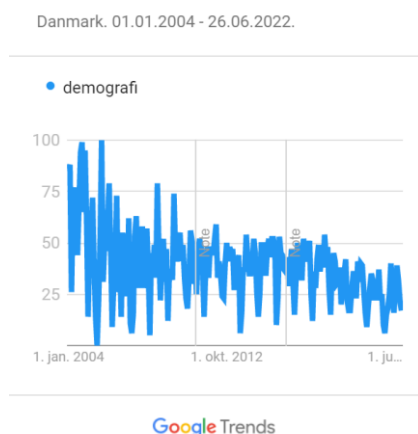
In the Nordic Statistical Database, table POPU07, population and population changes are presented in one Statbank table for all 8 countries. As the table has a correction component it equals what will be referred to as the Population Account's core table in this paper, without sex, age and birthplace. At aggregate level there are

absolutely no alarming issues, corrections are minor and leave the impression of world class statistics. But sex, age and cohort are not included in POPU07.

Demographic textbook calculations on demography cannot be fully replicated from real life free Nordic data. Many demographic indicators rely on sex and 1-year age/cohort to be estimated and if the information had been available mortality rates, life expectancy, survival measures could be documented, calculated and compared. The NSIs could share methods, translating documentation to their own language benefitting from shared intellectual property.

If the Population Account is not found in the same Statbank, we could vision to have the same table-structure published in each of our Statbanks. Standardizing the table structure will minimize the data collection process and enable us to learn from each other, even allowing for shared methods. By summer 2022, Hagstova Faroe Islands, and Statistics Greenland hold similar core tables in their Statbanks.

Takeaway 3: Vision a standard table structure

Age and cohort

Both age and cohort/year of birth are essential demographic characteristics and it is surprising to find the information partly being neglected in the Nordic national Statbanks. Only age at event is shown in published tables. Studies on cohorts need data from other sources. Studies based on microdata will miss the frame of knowledge which general Statbank tables offer to compare, check and understand results.

Survival indicators can be (and are) calculated directly from microdata, but results are not replicable, as microdata are generally not accessible.

Table 1. Availability in Official Nordic Statbanks

	Sweden	Denmark	Finland	Norway	Iceland	Greenland	Faroe Islands	Åland
age at event & age at end of year								
Population January 1st	✓	✓	✓	✓	✓	✓	✓	✓
Births	✗	✗	✗	✗	✗	✓	✓	✗
Death	✗	✗	✗	✗	✗	✓	✓	✗
Migrations	✗	✗	✗	✗	✗	✓	✓	✗
age at event								
Popuation January 1st	✓	✓	✓	✓	✓	✓	✓	✓
Births	✓	✓	✓	✓	✓	✓	✓	✓
Death	✓	✓	✓	✓	✓	✓	✓	✓
Migrations	✓	✓	✓	✓	✓	✓	✓	✓
age at end of year								
Population January 1st	✓	✓	✓	✓	✓	✓	✓	✓
Births	✗	✗	✗	✗	✗	✓	✗	✗
Death	✓	✗	✗	✗	✗	✓	✓	✗
Migrations	✗	✗	✗	✗	✗	✓	✗	✗

In a diagram, known as a Lexis diagram, the three dimensions are depicted as in Chart 1. When only age/period is available, it is only possible to estimate cohorts.

Chart 1. Lexis Diagram


Source: https://github.com/ottlngr/LexisPlotR
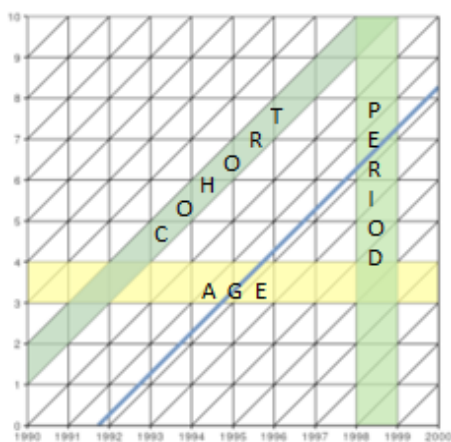
To summarize: only for Greenland and Faroe Islands emigration rates can be calculated by cohort, and only for Sweden, Greenland and Faroe Islands death rates by cohort can be calculated.

For all Nordics, rates can only be calculated by Lexis squares, a unit that consists of 2 consecutive cohorts. Lexis squares are perhaps just fine for larger populations where consecutive cohorts do not differ much in size. However, without data, we will not know. Neither can overall quality be scrutinized.

Proposed storage format with age and cohort

The Lexis diagram solves depicting the 3 Age-Period-Cohort dimensions in a 2-dimensional panel and likewise one has to be creative to store the same information in a data-cube, as the 3 dimensions in a cube will consist of mainly empty cells. In a calendar year, only 1 of 100 cohorts or 2 out of 100 ages will hold a figure, the rest will be zeros.

Chart 2. Lexis triangles
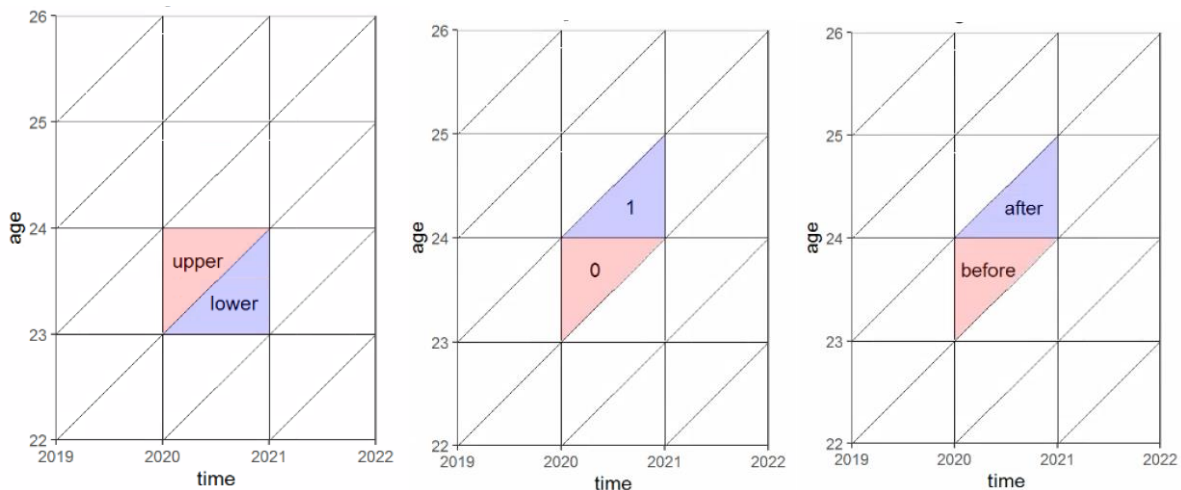
Lexis triangles are defined for flow components and of the three dimensions only cohorts and periods are stored in an event-table, where a variable called "Triangles(lexis)" is used to store information on whether events happened before (upper=0) or after birthday (lower=1). Age at time of event can now be calculated. This minimizes the number of empty cells.

Triangles are only defined for deaths, immigrations, and emigrations, whereas the other dimensions are equal to the ones found in the core table, below.

Chart 3a. Lexis events
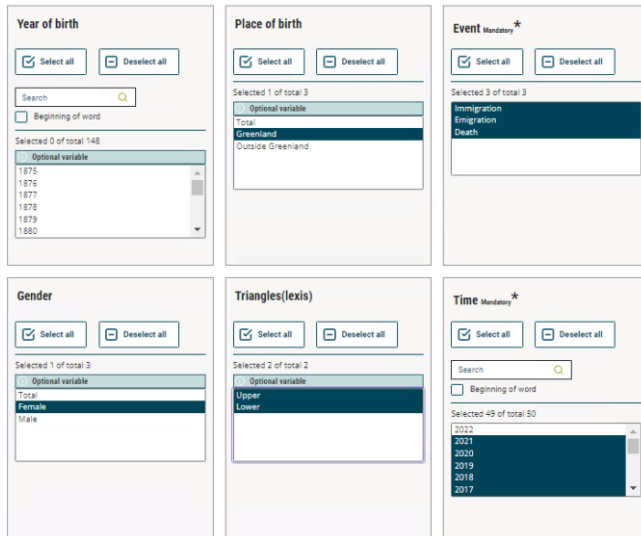
To keep it simple and for dissemination purposes the core table holds both population estimates and events, but no information on triangles, as they are only relevant for events.

Chart 3b. Core table
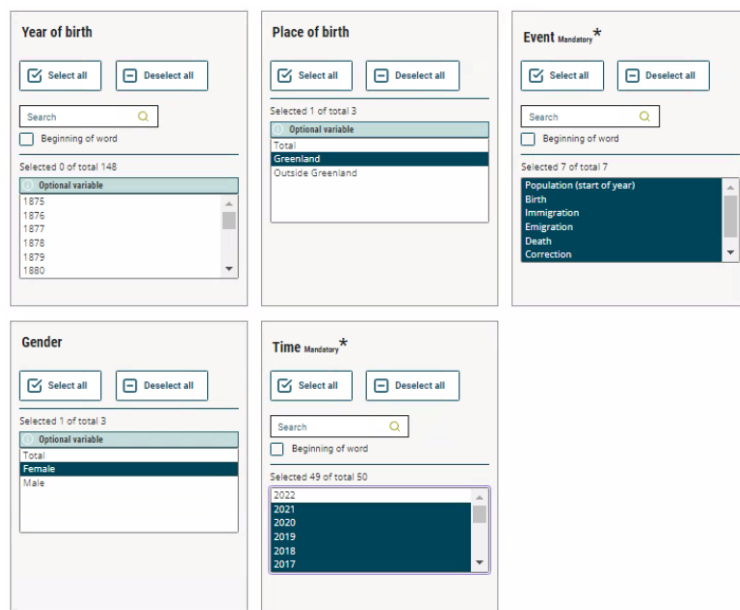
Source: Statbank Greenland

The core table combined with the Lexis event table supports all possible definitions of methods to calculate age specific death- and emigration rates. And all demographic key figures; life expectancy, age specific deathrate(s) can be calculated from the account.

The core table can be extended with additional dimensions, dependent on relevance and data availability.

To calculate fertility measures a satellite table with number of liveborn children by age of mother at birth and at end of year (with many zeroes) has been added.

To account for internal migrations another satellite table is added, holding counts of relocations (movings) between regions.

In addition, the Regional Population Account has two satellite tables, a) liveborn children and b) movings between regions

Chart 4. Regional Population Account by number of inhabitants. > 6,000
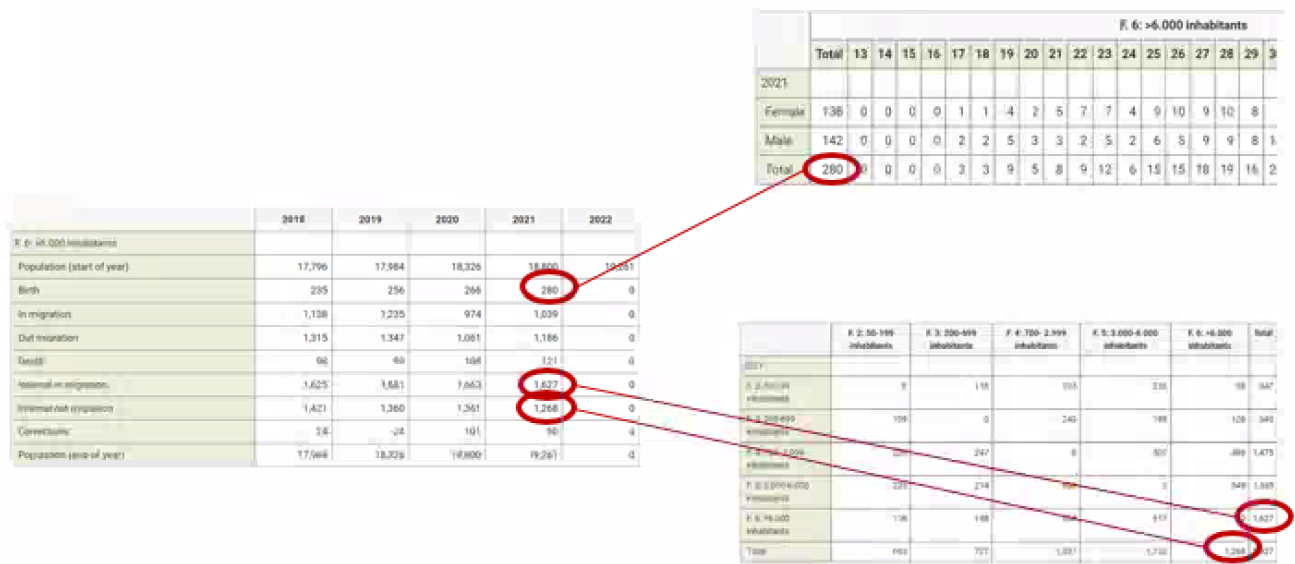

Takeaway 4: Adopt the envisioned Population Account structure

Case: Quality assessment

Corrections or residuals in the Population Account give valuable information on the quality of the statistics. This is not seen at aggregate level in the Nordic POPU07 table but only revealed when cohorts are added to the equation. The Greenlandic Population Account also adds sex and place of birth as central classification dimensions, not to mention, lower-level administrative regions.

Chart 5. Greenlandic corrections 1994-2021 by cohort and time

a)   b)   c)
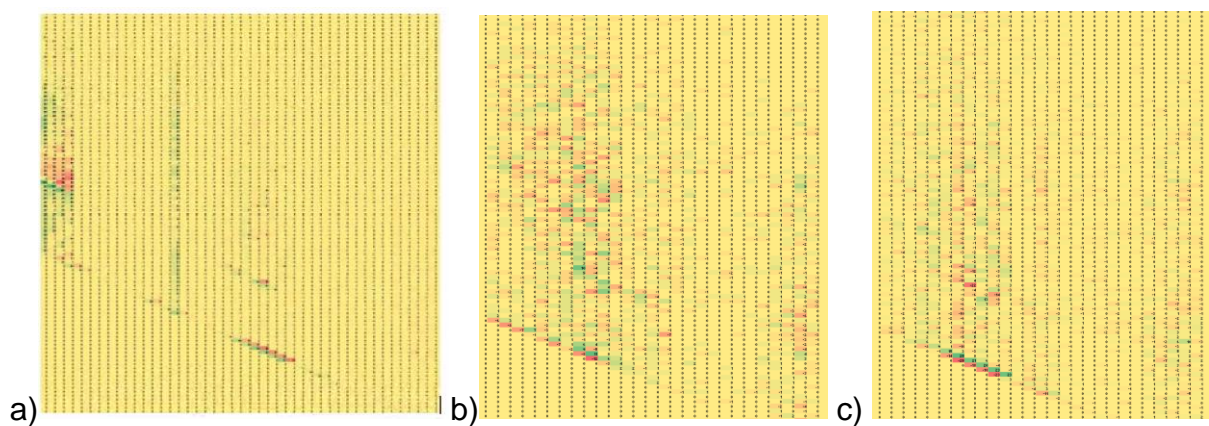
a) All Greenland 1973-2021, b) localities with less than 200 persons and c) localities with localities 700 – 3.000 persons

As illustration Chart 5 shows a simple visualization of the magnitude of corrections. It is a color-coded spreadsheet, green areas showing relatively large positive corrections, red relatively large negative corrections. And yes, both colors are statistical trouble zones, that needs inspection and actions to improve. A similar visualization on the Faroese population account shows another kind of trouble as corrections all count to zero. Data have been manipulated, as for sure also Faroe Islands too, must see late reports of either migration and/or deaths. This analysis is not possible from the other Nordic NSIs.
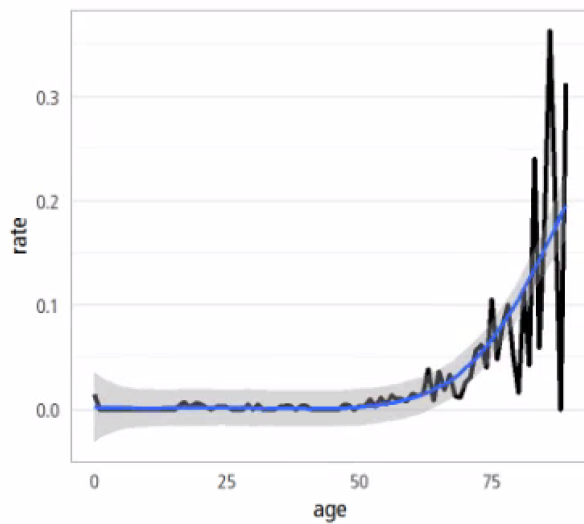
Takeaway 5: Do quality assessment

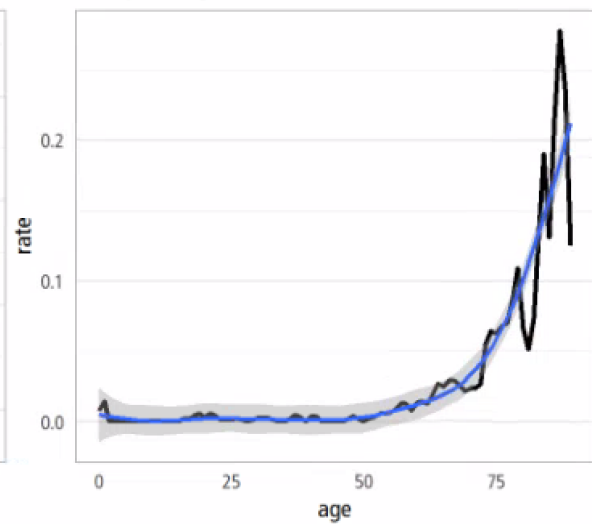**Case: Does age or cohorts matter? Greenlandic emigration and mortality rates**

In the Greenlandic case a relatively large correction post is seen when calculation base is Age (Lexis squares) but is not seen when base is Cohorts (Lexis parallelograms).

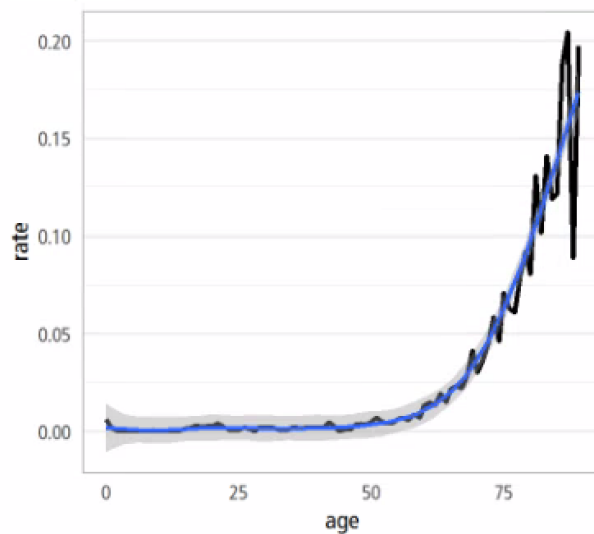## Chart 6. Mortality by age. 1 and 5-year base data compared
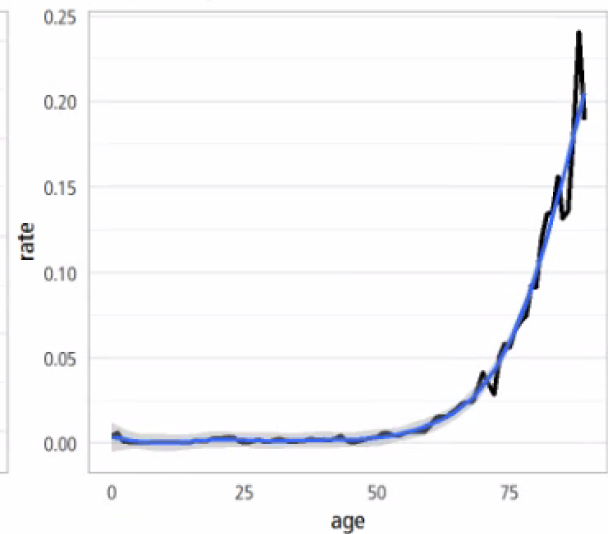
In Chart 6 above, 2*2 panels show mortality rates resp. for a 1-year and a 5-year period, for females born in Greenland. Calculated on one year base the Lexis-squares show much more volatility than when calculated on cohorts (parallelograms), as squares consists of 2 consecutive cohorts. Even when calculated for a 5-year period the cohort calculated rates are less influenced by corrections and thus les volatile, than square based calculations.

At age 16, a large proportion of Greenlanders move to Denmark for continuation school. For the 5-year period, 2015-2019, emigration rate can be calculated as either 19 or 25 pct depending on being calculated on age or cohort. Perhaps size matters, but the calculations cannot be carried out, without data available.

The full calculations can be followed in Appendix A and Appendix B and for any other combination filtered at https://larsp.shinyapps.io/lexis_greenland/

Takeaway 5: Ready to run documentation

Case: Life tables

Life expectancy at birth, or how many years a new-born baby can expect to live, is and has been the central health indicator used to compare countries and a society through centuries. Life tables are studied in many intellectual disciplines like medicine, actuarial science, economy, demography, etc.

The mortality calculations shown in Appendix A and B are self-documented and the code is used to generate a ready-to-publish Statbank-table, presenting mortality rates on all combinations of calculations whether it is Lexis squares or parallelograms, on 1-, 2- or 5-year base. From this, Lifetables can be calculated.

In his Ph.D. Thesis from 2018, Marius D. Pascariu studied 'Modelling and forecasting mortality' in depth. His theoretical studies are the base for his r-package 'Mortalitylaws', freely available for anyone to use. Functions convert age-specific mortality rates to all derived measures.

Some few lines of code is all it takes to convert age-specific mortality rates to a life tables

```
library(tidyverse)
library(MortalityLaws)

data_raw <- read_csv("https://bank.stat.gl/sq/c5c9592c-223f-4cb2-9ce6-461478c00e7
locale = locale(encoding = "latin1")) %>%
  filter(age<90) %>%
  mutate(rates=as.numeric(Mortalityrates)) %>%
  select(age,rates)

x <- 0:89
radix <- 1000
dataframe <- LifeTable(x, qx = data_raw$rates, lx0 = radix)
head(dataframe$lt,5)
```
`# R Markdown ⇕`

Terminal × Jobs ×

4.1.3 · S:/STATBEFO/GS Produktion/6. Fremskrivning/2022/Fremskrivning/Do_PX/LT/

```
nt x          mx       qx        ax        lx        dx        Lx        Tx        ex
1) 0 0.011040725 0.01098 0.4990799 1000.0000 10.980000 994.4999 67647.72 67.64772
2) 1 0.001821658 0.00182 0.4998482  989.0200  1.800016 988.1197 66653.22 67.39320
3) 2 0.002412909 0.00241 0.4997989  987.2200  2.379200 986.0299 65665.10 66.51517
4) 3 0.001811640 0.00181 0.4998490  984.8408  1.782562 983.9492 64679.07 65.67465
5) 4 0.000000000 0.00000 0.4998490  983.0582  0.000000 983.0582 63695.12 64.79283
```

In this case the Greenlandic Population Account serves as a data provider for the Life Table engine. In return the Mortalitylaws R-package delivers rich and detailed documentation, to which Statistics Greenland links. No need to prioritize publishing additional text, neither in Danish nor in Greenlandic.

Takeaway 6: Reuse results from researchers

Case: Demographic Research - Max Planck Institute

Based in Rostock, Germany, the Max Planck Institute for Demographic Research (MaxPlanck) collaborates with universities and researchers worldwide. Max Planck offers a number of standardized data collections from more than 40 countries on mortality (The Human mortality database (HMD) mortality.org) and fertility (HFD: humanfertility.org). Calculation methods are exemplarily described and published for free on these web resources. In addition, these are supported by R-packages that perform the calculations.

See https://www.mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf

The data collections allow researchers to benefit from a standardized access to mortality data. For comparative studies, standardization is pivotal. Researchers can focus on methods and analysis, spending as little time as possible on data wrangling.

A corresponding data resource is also in operation on Human fertility (HFD). But that is the limit. No migration, international or internal, no subnational data (except for US and Canada)

Data collections are updated annually, pushed by an internal operation process in the delivering NSI. As all data required are stored in the Greenlandic Population Account collecting data could be a pull process initiated by the receiving party, with no specific interference from the delivering party.

But Greenland is too small to be included in the data collections. MaxPlanck fears their methods aren't applicable to populations at sizes of 56,000. Nevertheless, researchers on Greenlandic demographic data might need data in the same format that can be extracted from mortality.org to compare mortality findings with Yukon and Northwest-Territories / Nunavut. To accomplish this, R-code can be supplied, to extract Greenlandic data and convert to MaxPlanck HMD formats

Takeaway 7: Integrate work and results of external researchers

Case: Projections

For population projections Statistics Greenland aim to share projection results from calculations based on data from the extended Population Account only. When ready, the model itself will be uploaded to GitHub as version controlled open source, coded in the free statistical analysis program, R.

Projecting future population, from one year to the next, is simple math, that needs to be defined only once, in a calculation module with output being saved to a Statbank table. The calculations depend on separate projections of the demographic components. Each component will store results in temporary dataframes with a uniform standard format.

This allows for alternate projections of each component from which the calculation module can pick.
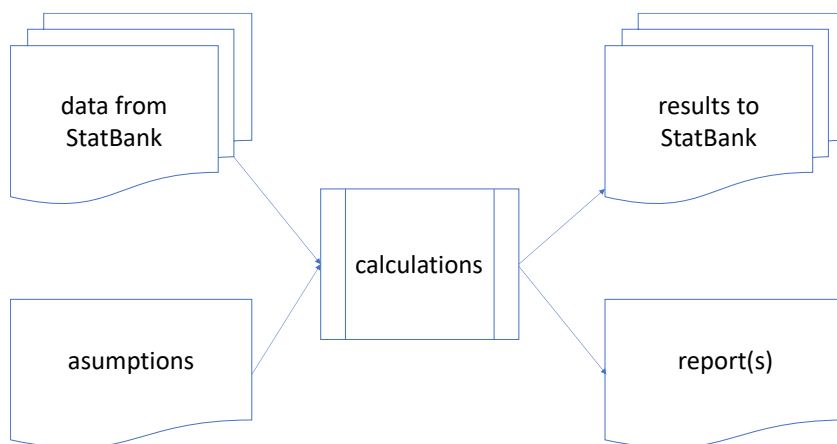
This approach offers full transparency, allowing anyone to get a copy of the model and scrutinize the calculations. This approach opens for endless alternate projections, which calls for standard reports with comparisons to the default projection results, found in https://bank.stat.gl
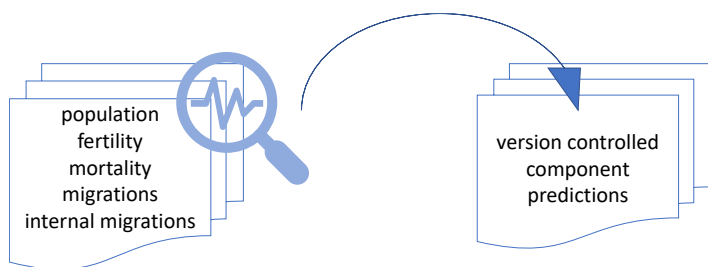
Takeaway 8: Do not think of the Statbank, as a teaser.

Early July 2022, Statistics Greenland and Hagstova Faroe Islands has started a co-operation with this described model as base. The 2 NSIs both use the file-based version of PxWeb and thus the same api. First step is to co-ordinate codes used in the Population Account tables to minimize differences and define a common projection core, that outputs px-files.

Mortality rates and Lifetables will be calculated from both Population Accounts as documented in Appendix A & B.

Analysis on country specific population data that generates projected components cannot be expected to be identical for our 2 countries. But we can surely inspire each other.

Open for business, not for sale

Disclaimer: content not approved by Statistics Greenland

Takeaway 9: The Statbank is an important resource for real life decisions

It is not the pure number of tables in a Statbank that secures relevance. Popular name statistics only attract attention, but do not deliver real value or insights, nor does statistics on 'average Greenlander', SDG indicators and the like, without in depth detailed statistics. To stay relevant our Statbank must give access to hard facts, at detailed level for public administration, businesses, and researchers to use and trust.

Tables must reflect society and serve as a model of our societies. Besides being timely it must give a coherent image to be credible and stay relevant.

The Population Account is the necessary set of tables that proves coherency and assess uncertainty. Exposing free and detailed data attracts researchers quid pro quo.

EUROPE   Published August 16, 2019 10:53am EDT
Greenland says it's 'open for business, not for sale' after Trump purchase report

POLITICS
Greenland Tells Trump: 'We're Open for Business, Not for Sale'
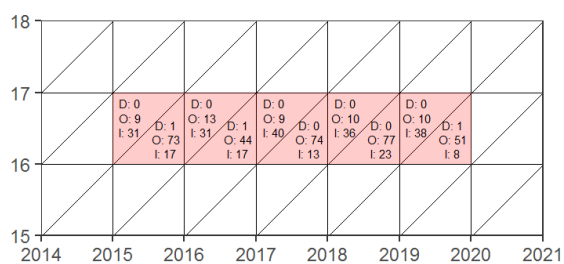Danish leaders dismiss President Trump's interest in purchasing the island; 'It must be an April Fool's Day joke'

'Open for business, not for sale': Greenland tells Donald Trump

No thanks, we're not for sale, aghast Greenland tells Trump

Greenland tells Trump it is open for business but not for sale

Greenland open for business with US, but not for sale — Foreign Ministry

Sum (red area) - Age: 16, Cohort: 1998 - 2003

time	D(Death)	I(Inmigration)	O(Outmigration)	Population(start)	Population(end)	Mean population	C(Corrections)
Total	3	254	370	1.497	1.440	1.468,5	62
2015	1	48	82	308	312	310,0	39
2016	1	48	57	312	285	298,5	-17
2017	0	53	83	285	289	287,0	34
2018	0	59	87	289	303	296,0	42
2019	1	46	61	303	251	277,0	-36

Deathrate (age: 16) = 3 / 1468.5 **= 0.0020429**

Emigration rate (age: 16) = 370 / 1468.5 **= 0.2519578**

Appendix B. Women Born in Greenland (2015 - 2019) – Lexis parallelograms
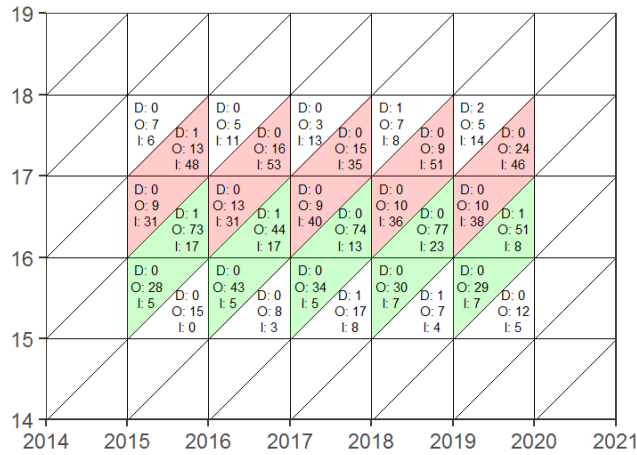

B1 (red area) - Age: 16/17, Cohort: 1998 - 2002

time	D(Death)	I(Inmigration)	O(Outmigration)	Population(start)	Population(end)	Mean Population	C(Corrections)
Total	1	409	128	1.497	1.786	1.641,5	9
2015	1	79	22	308	365	336,5	1
2016	0	84	29	312	367	339,5	0
2017	0	75	24	285	334	309,5	-2
2018	0	87	19	289	358	323,5	1
2019	0	84	34	303	362	332,5	9

B2 (green area) - Age: 15/16, Cohort: 1999 - 2003

time	D(Death)	I(Inmigration)	O(Outmigration)	Population(start)	Population(end)	Mean Population	C(Corrections)
Total	3	107	483	1.819	1.440	1.629,5	0
2015	1	22	101	392	312	352,0	0
2016	1	22	87	351	285	318,0	0
2017	0	18	108	379	289	334,0	0
2018	0	30	107	382	303	342,5	-2
2019	1	15	80	315	251	283,0	2

Deathrates (Death/Mean Population) calculated from totals in B1 & B2

src	Death	Population(start)	Population(end)	Mean Population	Deathrate
B1	1	1.497	1.786	1.641,5	0,0006091989
B2	3	1.819	1.440	1.629,5	0,001841056

Emigration rates (Emigration/Mean Population) calculated from totals in B1 & B2

src	Death	Population(start)	Population(end)	Mean Population	Deathrate
B1	128	1.497	1.786	1.641,5	0,07797746
B2	483	1.819	1.440	1.629,5	0,2964099

Mortality rate (age: 16) = sum(6.091989^{-4}, 0.0018411) / 2 **= 0.0012251**

Emigration rate (age: 16) = sum(0.0779775, 0.2964099) / 2 **= 0.1871937**