

Демографический ресурс для анализа и визуализации данных

Помазкин Д.В.¹, Филиппов В.А.¹

¹ АО «НПФ ГАЗФОНД», Москва, 117556, Россия

Получено 22 January 2022 ♦ Принято в печать 30 March 2022 ♦ Опубликовано 17 November 2022

Цитирование: DV Pomazkin, VA Filippov (2022) Demographic resource for data analysis and visualization. Population and Economics 6(3): 117–124. <https://doi.org/10.3897/popecon.6.e81027>

Аннотация

Демографический ресурс разработан для визуализации демографических данных в среде Интернет. В статье приведено краткое описание структуры базы данных и представлены примеры отчетных форм. При разработке данного ресурса использованы открытые источники демографической информации: данные The Human Mortality Database и данные Федеральной службы государственной статистики. Применение демографического ресурса открывает доступ к базовым демографическим характеристикам в разрезе регионов России с 1990 года. Помимо России и ее регионов данный ресурс содержит демографическую информацию по ряду стран, представленных в The Human Mortality Database. Разработанный двуязычный интерфейс (на русском и английском языках) позволяет формировать как табличные, так и графические отчетные формы – демографические возрастно-половые пирамиды с одногодичными возрастными группами, привычные динамические графики, тепловые карты изменения уровня смертности. Демографический ресурс можно использовать в учебных целях, при составлении аналитических материалов, отчетов и презентаций. Данные также можно выгрузить в формате csv для дальнейшего анализа и использования. Адрес ресурса: <http://demography.infoarchives.ru>

Ключевые слова

демографические базы данных, тепловые карты изменения уровня смертности, возрастно-половые пирамиды, визуализация

Коды JEL: C55

Введение

Анализ большого объема информации наиболее эффективен, когда отобранная информация представлена в компактном и удобном для обработки виде. Демографическая информация часто характеризует процессы, растянутые во времени, и представлена в виде временных рядов. Представление временных рядов в графическом виде существенно облегчает визуализацию данных и их анализ. Кроме того, часто возникает необходимость в представлении не одного числового ряда, а нескольких рядов для сравнительного анализа или определения дополнительных характеристик, например средневзвешенного значения и т.д. Аналогичные задачи возникают в различных областях, где требуется применение анализа временных рядов [Карминский, Помазкин, 1999; Карминский, Помазкин, 2000; Чернавский, 2001; Котляров и др., 2014].

Разработанный демографический ресурс дает возможность выгружать данные по заданным критериям, формировать как табличные, так и графические отчетные формы – демографические возрастно-половые пирамиды с одногодичными возрастными группами, привычные динамические графики, тепловые карты изменения уровня смертности.

Аналогичные решения представлены на ряде платформ, как специализированных, так и включающих демографическую информацию, среди которых можно выделить следующие:

- демографическая статистика по странам ОЭСР (<https://stats.oecd.org/>), включая прогноз численности населения в разрезе 5-летних групп;
- открытые данные Всемирного банка (<https://databank.worldbank.org/home.aspx>);
- демографические базы данных на сайте Французского института демографических исследований, объединяющие различные показатели в области демографии (смертность, рождаемость, миграция и т.д.), а также социологии, экономики и политики (<https://datalab.site.ined.fr/en/demographic-contextual-data>);
- World Factbook (<https://www.cia.gov/the-world-factbook/countries>).

Источники и структура данных

На демографическом ресурсе представлены годовые данные. Данные по регионам РФ предоставлены Федеральной службой государственной статистики (Росстат), они ежегодно дополняются, как правило, в августе следующего года. Данные в разрезе регионов РФ не корректировались нами по результатам переписей. Административно-территориальное деление не приводится к текущему состоянию, существовавшие ранее субъекты включены в базу данных.

Информация по странам мира получена путем экспорта данных, размещенных на сайте The Human Mortality Database (HMD) www.mortality.org, и ежегодно обновляется, а также присутствуют некоторые данные по странам, не представленным в базе HMD, полученные в результате выполнения отдельных проектов (они не обновляются). К таким странам относятся: Мексика, Саудовская Аравия, Корея, Турция и Узбекистан. Информация по данным странам включает половозрастное распределение и таблицы смертности в структуре, используемой для HMD. Включение данных стран преследовало цель сохранить и предоставить доступ к дополнительной информации и продемонстрировать возможности системы по расширению ее функционала.

Данные хранятся в реляционной базе данных, которая содержит следующие поля в разрезе однолетних половозрастных групп:

- год наблюдения,

- возраст, лет (от 0 до 110 лет),
- пол (мужской, женский),
- численность населения, тысяч человек,
- число проживающих до возраста x (корень таблицы дожития 100 000),
- число умерших, человек.

Данные представлены для российских регионов с 1990 г. (Севастополь с 2015 г.), для стран — в соответствии с имеющейся информацией в базе данных HMD (например, для Франции с 1816 г.).

Функционал

Демографический ресурс состоит из двух основных разделов “Некоторые страны мира” и “Регионы России”, внутри каждого из них есть следующие вложения:

- **Общие данные**
- **Возрастная структура**
- **Теплокарты смертности** (только для раздела по странам мира).

В разделе «Общие данные» можно отобразить и сравнить данные нескольких стран или регионов России по таким показателям:

- численность населения, тысяч человек,
- ожидаемая продолжительность жизни в возрасте x , лет,
- ожидаемая продолжительность жизни городского населения в возрасте x , лет,
- доля населения в выбранном интервале возрастов, процент,
- нагрузка пожилыми людьми (отношение численности населения в возрасте старше 60 лет к численности населения в возрасте 20-59 лет),
- число умерших, человек.

Интерфейс также содержит фильтры для выбора периода представления данных, возраста и пола рассматриваемой группы населения. Ниже приведены примеры выборки и визуализации данных раздела «Общие данные» (рис. 1–3).

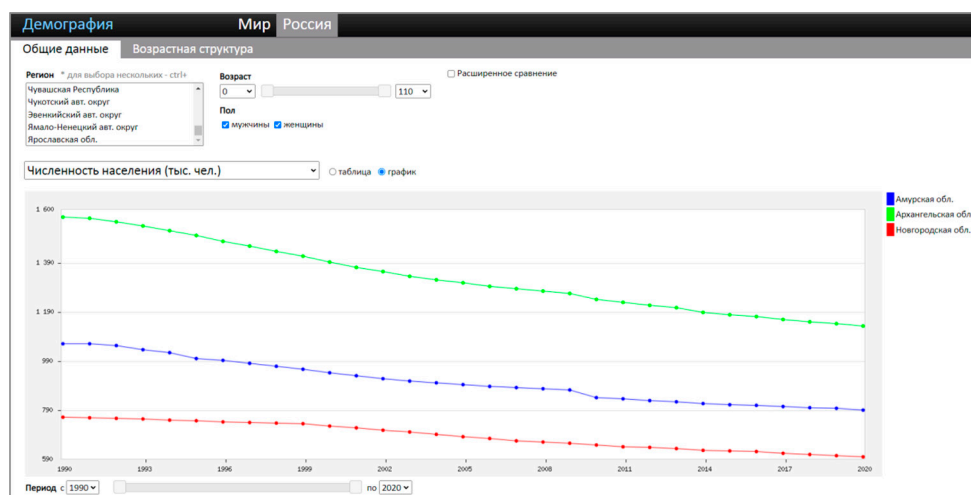


Рисунок 1. Численность населения Амурской, Архангельской и Нижегородской областей за период с 1990 по 2020 годы, тысяч человек

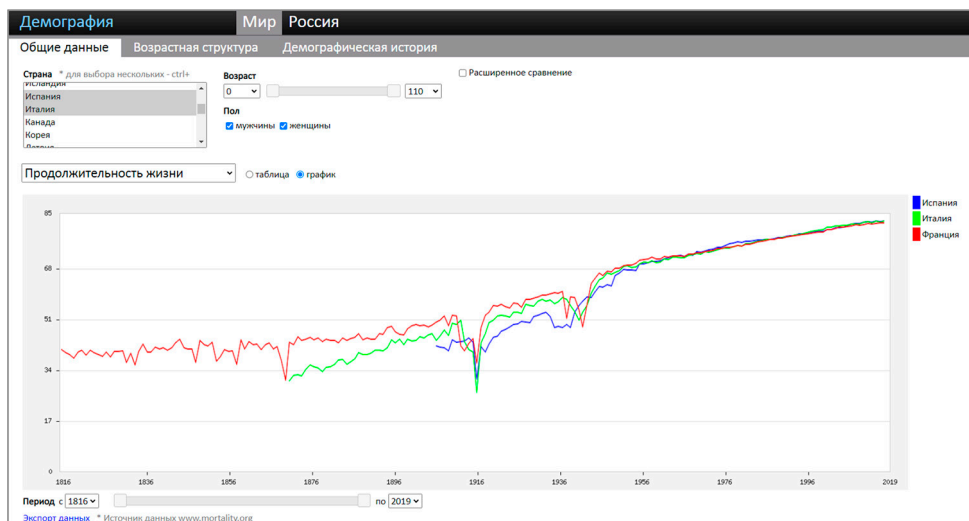


Рисунок 2. Ожидаемая продолжительность жизни при рождении населения Испании, Италии и Франции за период с 1816 по 2019 годы, лет

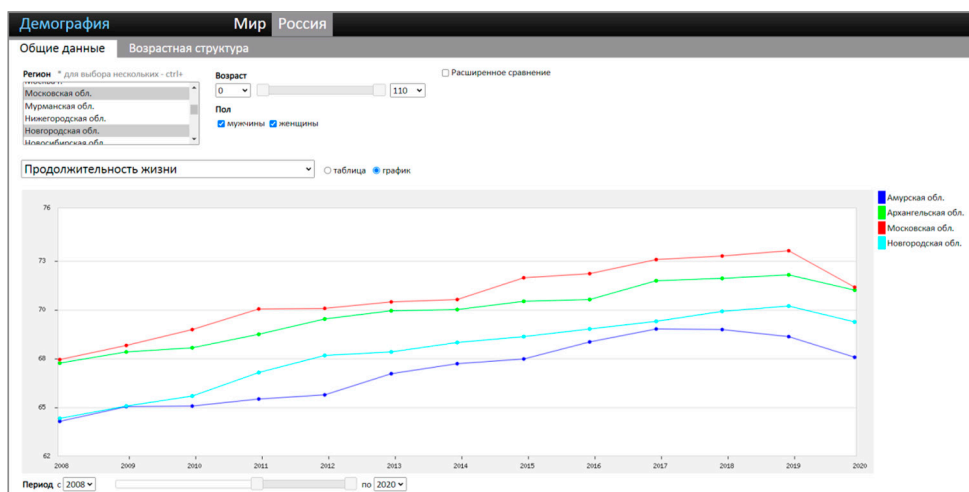


Рисунок 3. Ожидаемая продолжительность жизни при рождении населения Амурской, Архангельской, Московской и Новгородской областей за период с 2008 по 2020 годы, лет

В разделе «**Возрастная структура**» в графическом виде отображается диаграмма возраст-половой структуры населения с возможностью просмотра ее изменения по годам. Для этого используется слайдер, который при перемещении интерактивно отображает динамику изменения возрастной структуры. Пример выгрузки и визуализации данных по возраст-половой структуре приведен на рис. 4.

В разделе «**Теплокарты смертности**» ресурс отображает теплокарту ежегодного изменения уровня смертности для выбранной страны за указанный период времени. Двумерная визуализация наглядно отображает всплески заболеваний, войны и прочие явления, приводящие к изменению уровня смертности. Подобное представление данных позволяет оценить измене-

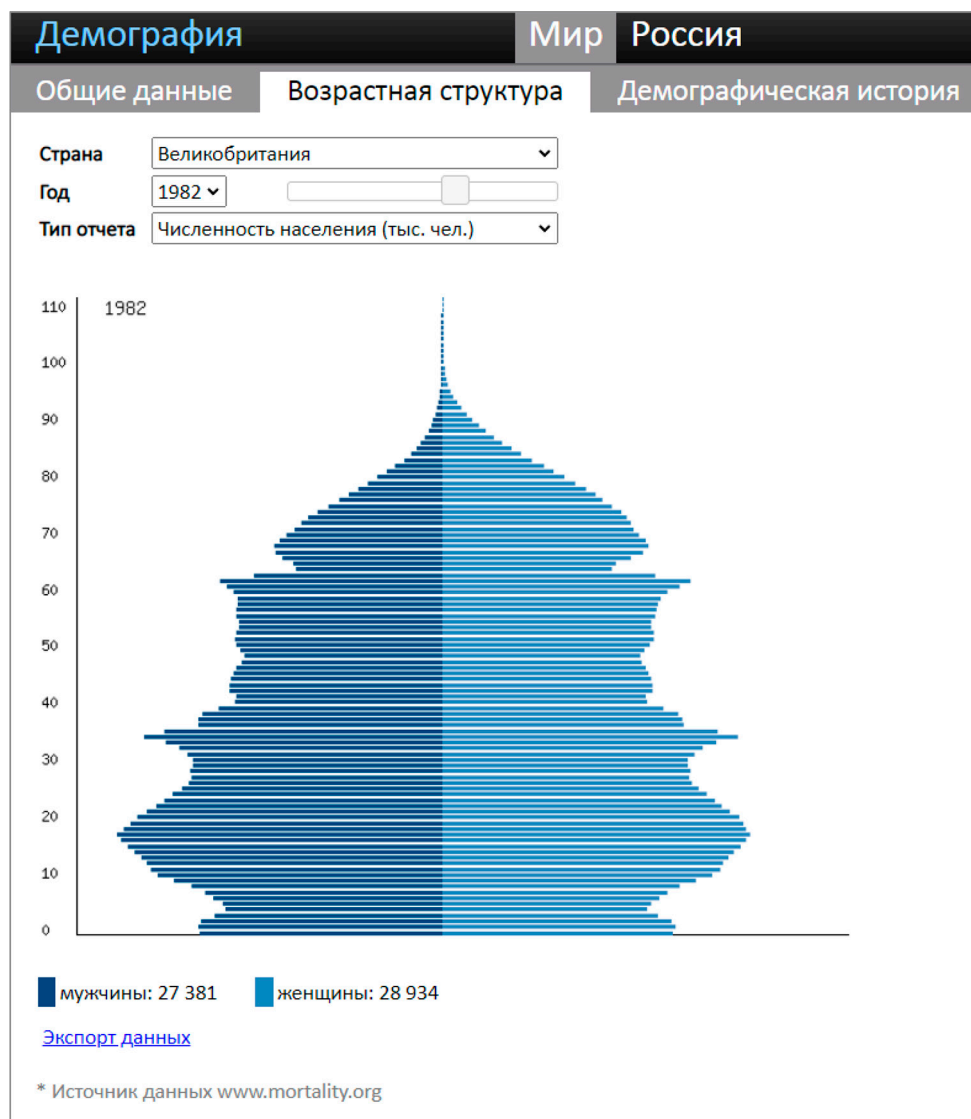


Рисунок 4. Возрастно-половая структура населения Великобритании в 1982 году

ние коэффициентов смертности в течение времени и выделить возрастные группы, наиболее сильно подверженные воздействию. Просматриваемая зависимость между историческими событиями позволяет заключить о сохранении истории в новом цифровом формате, носителем которого и являются таблицы смертности.

В настройках теплокарты можно управлять уровнем шума, изменяя отражаемую амплитуду уровня смертности, коэффициентами масштабирования осей X и Y и цветовыми каналами RGB.

Пример выгрузки и визуализации данных по тепловой карте смертности представлен на рис. 5.

Общие данные

Возрастная структура

Теплокарты смертности

Страна

Франция

Период

1900

-

1950

Возраст

0

-

100

Пол

мужчины

NX/NY

8

/

4

Шум

10

☒ RGB G-channel

Тип отчета

☐ таблица

☒ карта (ежегодное изменение уровня смертности в (%))

[Посчитать](#) (* [описание процедуры расчета](#))

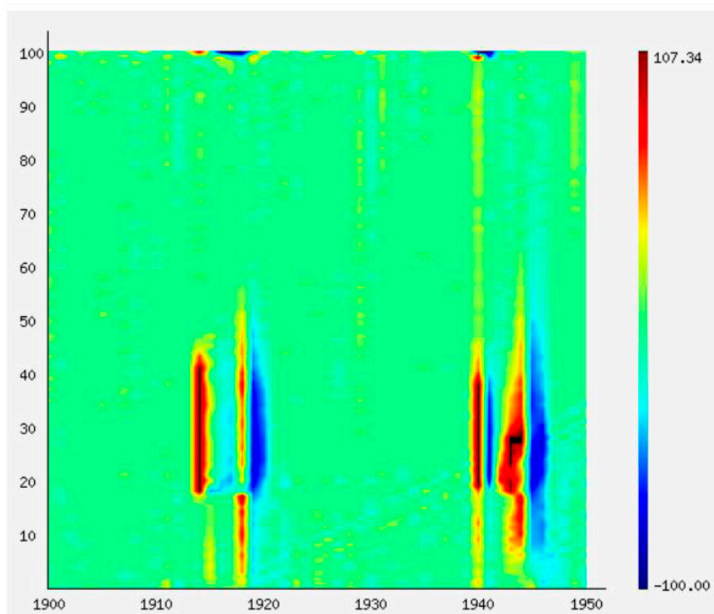


Рисунок 5. Тепловая карта смертности населения Франции за период с 1900 по 1950 годы

Метод построения тепловых карт смертности

Исторические данные о смертности свидетельствуют о сильной зависимости коэффициента смертности не только от возраста, но и от времени. Поэтому изучение явлений изменения уровня смертности является важной задачей.

Рассмотрим способ, позволяющий визуализировать изменение уровня смертности в течение времени в каждом возрасте. Для этого сопоставим вероятности смерти в каждом возрасте в течение заданного промежутка времени.

Вероятность умереть в возрасте (a) в год (t) получим как:

$$\mu(x, s, t) = \left(1 - \frac{l_{x+1}(s, t)}{l_x(s, t)} \right),$$

где $l_x(s, t)$ – число лиц пола (s), доживших до возраста (x), согласно таблице дожития в год (t).

Изменение вероятности смертности за год определим следующим образом:

$$\Delta\mu(x, s, t + 1) = \frac{\mu(x, s, t + 1)}{\mu(x, s, t)}.$$

Далее, проецируя полученное значение, выраженное в процентах, на плоскость время – возраст, получаем теплокарту для определенного пола. Вертикальная шкала – это проценты изменения вероятности смерти.

На рис. 6 такое поле смертности приведено для российских женщин в течение 1960–2006 гг. По вертикальной оси отложен возраст, по горизонтальной – время. В данном примере черный цвет соответствует росту вероятности смерти, белый – снижению. Цвет фона распространяется на ежегодные процентные колебания смертности в диапазоне $\pm 5\%$, что соответствует среднему значению положительных и отрицательных значений процентных изменений вероятностей смерти. Треугольниками отмечена линия ожидаемой продолжительности жизни при рождении в каждом календарном году.

Подобное представление данных позволяет оценить изменение коэффициентов смертности в течение времени и выделить возрастные группы, наиболее сильно подверженные данным изменениям. На рис. 6 мы видим рост смертности у женщин в России в первой половине 1990-х годов.

Более подробно с результатами подобных исследований можно ознакомиться в работах, посвященных демографической истории [Тимофеев, Помазкин].

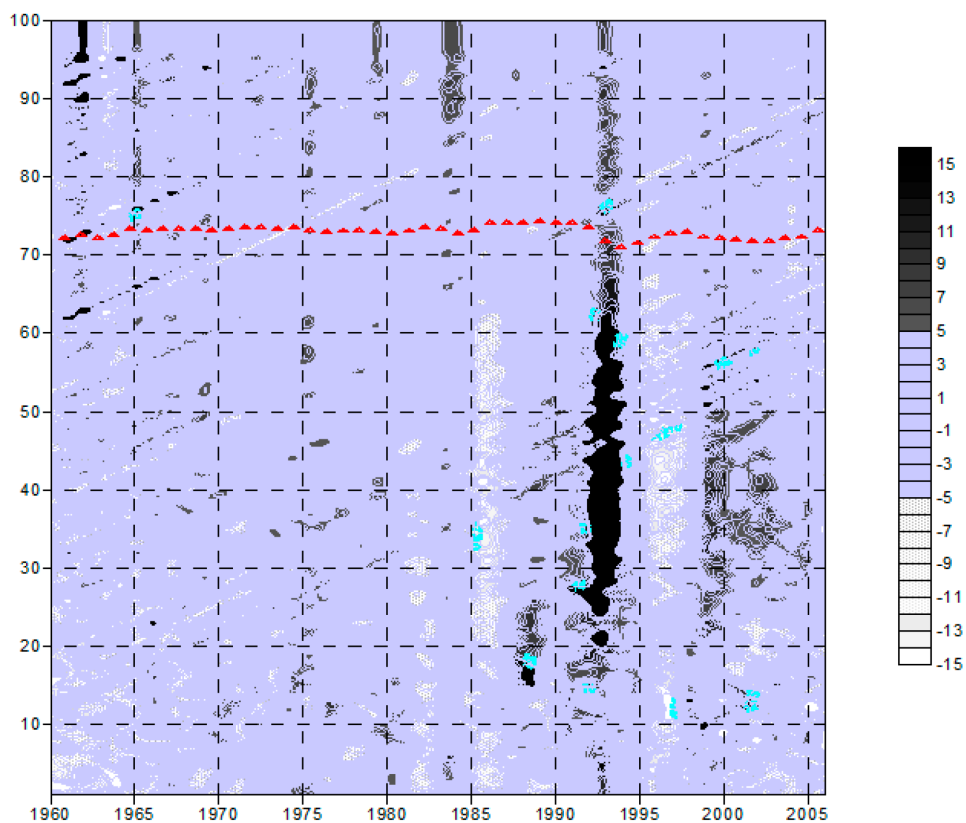


Рисунок 6. Тепловая карта ежегодных изменений вероятности смерти женщин в России в 1960-2006 годах. *Источник:* таблицы смертности за период 1960-2007, НМД

Заключение и перспективы развития ресурса

В распоряжение читателей представлен демографический ресурс по анализу и визуализации данных, полученных на основе открытых источников Федеральной службы государственной статистики и The Human Mortality Database, который позволяет формировать табличные и графические отчетные формы – демографические возрастно-половые пирамиды с одногодичными возрастными группами, привычные динамические графики, тепловые карты изменения уровня смертности. Демографический ресурс обладает двуязычным пользовательским интерфейсом (русский и английский) и может быть использован в учебных целях, при составлении аналитических материалов, отчетов и презентаций; кроме того, данные выгружаются при необходимости в формате csv для дальнейшего анализа и использования.

В ближайшее время планируются работы по автоматизации подготовки и загрузки данных, адаптации приложения под мобильные устройства и расширению функционала экспорта данных.

Благодарности

Авторы выражают благодарность Первому вице-президенту АО «НПФ ГАЗФОНД» В.В. Потехиной за оказанную помощь при выполнении работы.

Список литературы

- Карминский А.М., Помазкин Д.В. (1999) Графический анализ данных в Интернете // Банковские технологии: 2: 66-7.
- Карминский А.М., Помазкин Д.В. (2000) Образное представление данных // Рынок ценных бумаг: 4(163): 77-9.
- Котляров О.Л., Помазкин Д.В., Филиппов В.А. (2014) Контроллинг в сфере ИТ // С.Г. Фалько (ред.) Современное предприятие и будущее России. Сборник научных трудов международного форума, посвященного 85-летию кафедры «Экономика и организация производства» МГТУ имени Н.Э. Баумана. М.: НП «Объединение контроллеров», 130-5. URL: <http://controlling.ru/files/65.pdf>
- Чернавский Д.С. (2001) Синергетика и информация. Динамическая теория хаоса. М.: Наука.

Другие источники информации

- Тимофеев С.А., Помазкин Д.В. Демографическая история. URL: <http://infoarchives.ru/old-site/data/di.pdf>
- ФСГС (Росстат). URL: <https://rosstat.gov.ru/>
- Human Mortality Database. URL: www.mortality.org

Сведения об авторах

- Помазкин Дмитрий Владимирович, кандидат экономических наук, негосударственный пенсионный фонд «АО ГАЗФОНД», начальник управления. E-mail: D.Pomazkin@gazfond.ru
- Филиппов Владимир Андреевич, негосударственный пенсионный фонд «АО ГАЗФОНД», начальник отдела. E-mail: V.Filippov@gazfond.ru