

УДК 004.022, 004.62, 004.65, 004.422

## АНАЛИЗ СТАТИСТИЧЕСКИХ ИНФОРМАЦИОННЫХ РЕСУРСОВ И ИНСТРУМЕНТОВ С ИСПОЛЬЗОВАНИЕМ ГИС-ТЕХНОЛОГИИ

Никонов Эдуард Германович, доктор физико-математических наук, профессор Института системного анализа и управления, ГБОУ ВО МО «Университет «Дубна»

Самсонов Алексей Дмитриевич, студент Института системного анализа и управления, ГБОУ ВО МО «Университет «Дубна»

### Аннотация

*В данной работе приведён анализ программных продуктов и инструментов с применением ГИС-технологии. Как показывает деятельность человека, технологии ГИС применяются очень активно в статистике, в задачах прогнозирования, мониторинга территории и др. Затраты, денежные и умственные, на реализации продуктов в данной отрасли высоки. Поэтому очень важно, чтоб сам готовый продукт не только выполнял необходимые действия, но и был удобным. Инструмент также играет немаловажную роль. От грамотно поставленной задачи выяснится, какой инструмент нужен. Также от грамотной проработки функционала продукта, и проработки сценариев взаимодействия пользователя с ним, зависит, как часто продукт нуждается в модернизации, и как он часто будет пользоваться клиентами. В работе будут показаны, по каким критериям производится оценка разработки продуктов с технологиями ГИС. И по данным критериям, в дальнейшем определится с инструментами. Был рассмотрен такой вид продукта как статистический информационный ресурс, где вопрос об удобстве пользования продуктом не так часто затрагивается. Данных по миру в различных отраслях – много, и очень важно предоставить возможность пользователю ими грамотно распорядиться по поставленной задаче. И также важно сформулировать требования продукта. Ибо задачи будут не только по выявлению и отображению самих сведений, но и дальнейшая отчётность и презентация.*

**КЛЮЧЕВЫЕ СЛОВА:** статистические информационные ресурсы, статистика, геоинформационные системы, пространственные данные, визуализация данных, большие данные.

## ANALYSIS OF STATISTICAL INFORMATION RESOURCES AND INSTRUMENTS USING GIS TECHNOLOGY

Nikonov Eduard Germanovich, Doctor of Science (Phys & Math), professor, Dubna State University, Institute of the system analysis and management

Samsonov Alexey Dmitrievich, student, Dubna State University, Institute of the system analysis and management

### Abstract

*This short review is devoted to the analysis of software products and tools using GIS technology. As human activity shows, GIS technologies are used very actively in statistics, in the tasks of forecasting, monitoring the territory, etc. The costs, monetary and mental, for the sale of products in this industry are high. Therefore, it is very important that the finished product itself not only performs the necessary actions, but also is convenient. The tool also plays an important role. From a correctly posed task, it turns out which tool is needed. In addition, from a competent study of the product's functionality, and the development of user interaction scenarios with it, it depends on how often the product needs to be modernized, and how often customers will use it. The review will show what criteria are used to evaluate the development of products with GIS technologies. In addition, according to these criteria, in the future to determine the tools. It was examined such a product as a statistical information resource, where the question of the convenience of using the product is not so often addressed. There are many data on the world in various industries, and it is very important to provide the user with the opportunity to correctly manage them according to the task. Moreover, it is important to formulate product requirements. For the tasks will be to not only identify and display the information itself, but also further reporting and presentation.*

**KEYWORDS:** statistical information resources, statistics, geographic information systems, spatial data, data visualization, big data.

## **Введение**

Работа с данными – один из важнейших этапов развития компании и общества в целом. Без работы с данными не были бы возможны выявления закономерностей и принципов, теорий и их доказательства. Работы с данными помогают науке, людям в их деятельности: от математиков до инженеров, от социологов до метеорологов и так далее. Поэтому важно иметь ресурсы, благодаря которым можно обработать свои данные, встроенные данные, и в дальнейшем выводить данные в определённом формате, нужный для эксперта по его задачам.

«Всегда, когда новая технология приживается, люди перестают обращать внимание на красочные рекламные обещания, появляется спрос на практичный и понятный дизайн. Тогда производители пересматривают дизайн и применяют к нему те же принципы, на которых был основан дизайн предыдущего поколения техники. Самые вопиющие ошибки допускают разработчики новейших технологий» [1].

В данной работе будут рассмотрены существующие аналитические решения по обработке данных, известные как статистические информационные ресурсы. Их функционал и недостатки, какие подходы есть в обработке и выводе данных. А также программные продукты с использованием технологии ГИС.

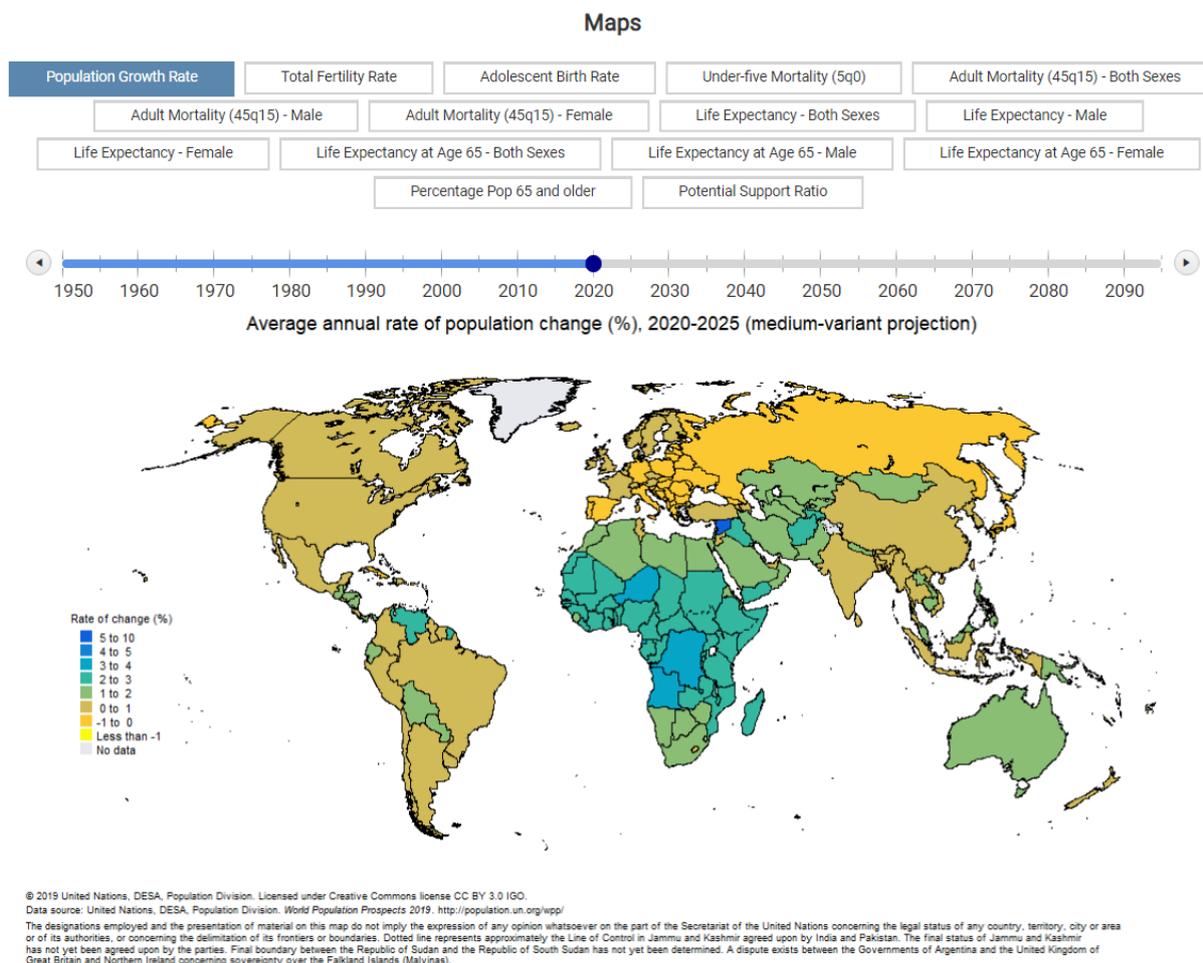
## **Статистические информационные ресурсы**

«Официальная статистика является необходимым элементом информационной системы демократического общества, обеспечивая правительство, экономические круги и общественность данными об экономическом, демографическом, социальном и экологическом положении» [2].

«Официальная статистическая информация является общедоступной, за исключением информации, доступ к которой ограничен федеральными законами. Обеспечение доступа заинтересованных пользователей официальной статистической информацией к общедоступной официальной статистической информации осуществляется путем ее предоставления или распространения» [3].

Статистический информационный ресурсы – частный случай информационного ресурса. Под информационным ресурсом понимается долговременное хранилище, способ тиражирования и передачи смыслов, выраженных человеко-читаемыми и машиночитаемыми знаками, целенаправленно созданное для использования в качестве ресурса/средства определенной деятельности и потому имеющее ярко выраженные потребительские ценности, которые зачастую позволяют позиционировать этот продукт в качестве товара на

информационном рынке [4, стр. 5]. В нашей работе «смыслом» через рассматриваемые знаки являются различные виды статистики: от таблиц до отчётов, до растровых изображений и так далее. Поэтому под статистическим информационным ресурсом будем понимать информационный ресурс по работе со статистическими данными, то есть не только с входящими данными, но и с данными после проведения статистического анализа (рис. 1).



**Рис. 1. Статистический информационный ресурс**

Рассмотрим следующие статистические информационные ресурсы:

- *FAOSTAT (Food and Agriculture Organization Corporate Statistical Database);*
- *ILOSTAT (International Labour Organization);*
- *UNSD (United Nations Statistic Division);*
- *IMF (International Monetary Fund);*
- *UNECE (United Nations Economic Commission for Europe).*

### **FAOSTAT**

*FAOSTAT* известно под именем «Продовольственная и сельскохозяйственная организация ООН». Ориентируется на уничтожение голода по всему миру. В задачах ресурса

– оснащение статистических сведений для снабжения продовольственной безопасности и обеспечения доступа населения к высококачественной пище (рис. 2).



Рис. 2. FAOSTAT

### ILOSTAT

*ILOSTAT* известно как «Международная организация труда». Оно предоставляет пользователям исчерпывающие международные данные по широкому кругу вопросов, связанных с трудовой деятельностью (рис. 3).



Рис. 3. ILOSTAT

### UNSD

*UNSD* известно как «Статистический отдел ООН». Оно собирает и распространяет глобальную статистическую информацию. Разрабатывает стандарты и нормы для статистической деятельности и поддерживают усилия стран по укреплению своих национальных статистических систем (рис. 4).



Рис. 4. UNSD

### IMF

*IMF* известно как «Международный валютный фонд». Оно контролирует политику стран-членов, а также национальные, региональные и глобальные экономические и финансовые события с помощью формальной системы, известной как надзор (рис. 5)



Рис. 5. IMF

## UNECE

*UNECE* известно под именем «Европейская экономическая комиссия ООН». Оно способствует повышению эффективности ООН посредством регионального осуществления решений глобальных конференций и встреч на высшем уровне ООН (рис. 6).



Рис. 6. UNECE

Каждый из рассматриваемых ресурсов имеют свои преимущества и недостатки. Были выбраны следующие критерии оценивания:

- экспорт изображения диаграмм;
- экспорт документов в виде диаграмм;
- экспорт данных;
- возможность самостоятельно делать диаграммы;
- интерактивный режим с ГИС.

В итоге была сформирована следующая таблица.

**Таблица 1. Критериальная оценка ресурсов**

	Экспорт изображения диаграмм	Экспорт документов в виде диаграмм	Экспорт данных	Возможность самостоятельно делать диаграммы	Интерактивный режим с ГИС
<i>FAOSTAT</i>	+	+	+	-	+
<i>ILOSTAT</i>	+	-	+	+	+
<i>UN</i>	-	+	+	-	+

	Экспорт изображения диаграмм	Экспорт документов в виде диаграмм	Экспорт данных	Возможность самостоятельно делать диаграммы	Интерактивный режим с ГИС
<i>IMF</i>	-	-	+	-	-
<i>UNECE</i>	+	-	+	-	+

Во время исследования рассматриваемых ресурсов были выявлены следующие недочёты разработки, можно их распределить по принципам удобного пользования продуктом:

- Принцип соответствия (то есть программный продукт делает те действия, которые пользователь не хотел, чтоб они были):
- *Открывание нового окна или страницы браузера при попытке получения таблицы с графика;*
- *Блокировка страницы при долгом отсутствии действий пользователя;*
- *Перелистывание страницы вместо выпадающего списка с графика.*
- Принцип обратной связи (то есть программа с запозданием выполняет действие, или не оповещает пользователя о том, что программа выполняет какие-то действия долго, или не выполняет действия вообще):
- *Не работал выход из полноэкранного режима карты по кнопке на панели;*
- *Нет оповещения пользователя о пределе укрупнения или уменьшения рассматриваемого участка планеты для статистики.*
- Отсутствие программных ограничителей (излишняя свобода действий для пользователя, из-за которых тот не способен вернуться к поставленной задаче):
- *Чрезмерное укрупнение или уменьшение карты;*
- *Можно «спрятать» карту;*
- *Можно отобразить неограниченное количество графиков на одной диаграмме.*

### **Оформление отчётов**

Важно, чтоб данные были не только введены и обработаны, но и были выведены в вид, который требует пользователь в роли эксперта. Он может и не знать о деталях составления грамотного отчёта, но ему можно подсказать о возможностях, которые он не ожидал увидеть в продукте.

Одна из таких возможностей звучит следующим образом: «*Reduce the non-data ink. Enhance the data ink*» [5]. Смысл заключается в том, чтоб визуализация данных были не только хорошо читабельными в цвете, но и в чёрно-белых тонах. Ибо пользователи для распечатывания отчётов используют чёрно-белые лазерные принтеры. И для глубокого,

большого чтения предпочитают использовать электронные книги, которые в чёрно-белых тонах, а не цветные мониторы или смартфоны, портящие глаза.

Зачастую цветные графики не приспособлены под отчёты чёрных тонов (рис. 7).



Рис. 7. Демонстрация несоблюдения принципа с тонами

При работе с растровыми данными необходимо также этот принцип учитывать при разработке статистических карт. Далее рассмотрим с теории графов некоторые правила при выполнении данной задачи. В данном случае – отображение случая с теоремы о 4 красках на примере штатов континентальных США (рис. 8).

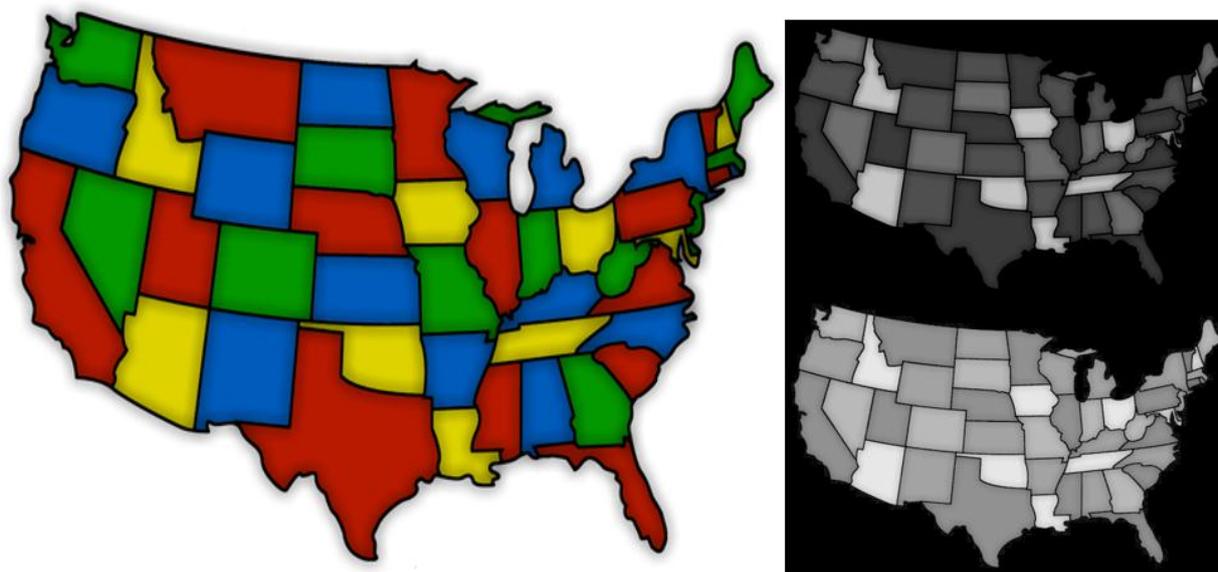
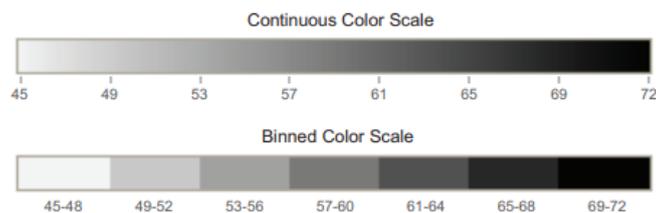


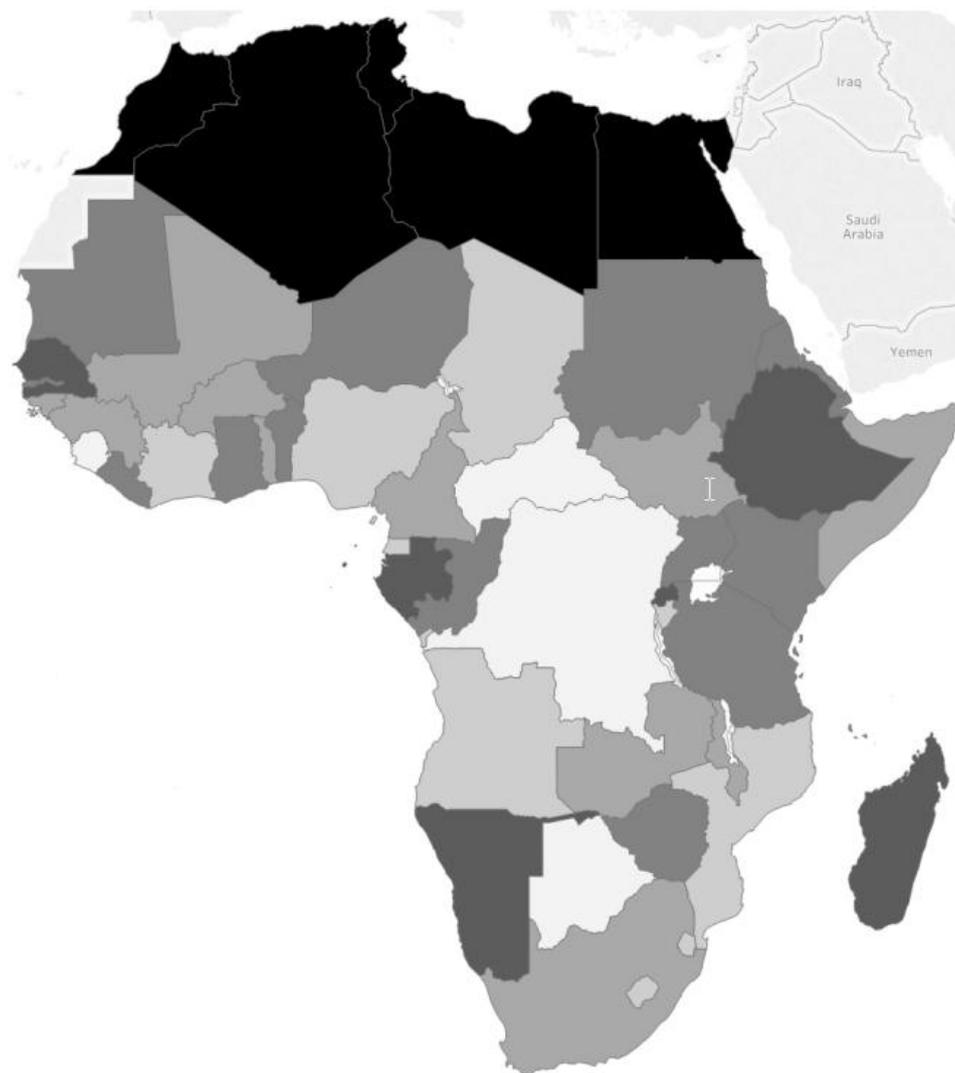
Рис. 8. Пример несоблюдения принципа с тонами на статистической карте

Была произведена попытка автором этого принципа о тонах отображения стран Африки по легенде чёрно-белого цвета [6]. В данном случае – работа со средней продолжительностью жизни мужского населения по странам. Как известно, есть шкала градиентная (*Continuous Color Scale*), а есть «по кускам» (*Binned Color Scale*), где в пределах данных значений – свой цвет (рис. 9). Первая требует много ресурсов для нахождения своего цвета, а второе соответственно – не требует.



**Рис. 9. Шкалы**

Автор обратил внимание, что цвета стран северной Африки совпадают. И издали покажется, что это одно большое государство, а не несколько государств поменьше (рис. 10).



**Рис. 10. Проблема отображения северных стран Африки**

Поэтому было предложено следующий вариант решения поставленной проблемы (рис. 11):

- Определить нижние и верхние границы рассматриваемых данных;
- Использовать масштаб для отображения данных по кускам;
- Внутри одного куска сделать «свою» градиентную шкалу (для различия одного участка от другого).



## Технологии ГИС

После рассмотрения проблем статистических информационных ресурсов и подачи выходных данных, можно браться за разработку своего программного продукта.

Были на рассмотрение выбраны следующие компоненты для разработки и их продукты:

- Языки программирования:
- *Python 3.7+*;
- *JavaScript*;
- *Java*.
- Серверы для развёртки ГИС:
- *MapServer*;
- *Geoserver*.
- Система управления базами данных:
- *MySQL*;
- *PostgreSQL*;
- *MongoDB*.
- Система управления контентом для ГИС:
- *Mapbender*.

Рассмотрим продукты каждого компонента по сравнительной таблице:

- Языки программирования;

**Таблица 2. Языки программирования**

	<i>Python</i>	<i>JavaScript</i>	<i>Java</i>
Пакетный менеджер	<i>PIP</i> (встроенный)	<i>NPM</i> (требует установки <i>NodeJS</i> )	Через среду разработки
Рейтинг <i>TIOBE</i>	3	7	1
Модификации для ведения статистики и расчётов	Есть	Нет	Нет
Коммерческая привлекательность	Средняя	Средняя	Высокая
Фреймворк для разработки веб-приложений	<i>Python Django</i>	<i>AngularJS</i>	<i>SpringMVC</i>

- Серверы для развёртки ГИС;

**Таблица 3. Серверы для развёртки ГИС**

	<i>MapServer</i>	<i>Geoserver</i>
Написан на языке программирования	<i>C</i>	<i>Java</i>
Наличие встроенного портала для скачивания	Есть	Нет

	<i>MapServer</i>	<i>Geoserver</i>
дополнений		
Внутреннее редактирование карты	Отсутствует	Присутствует
Ручной способ редактирования карты	Своя вариация языка (.map)	Основан на вариации XML
Поддержка современных стандартов картографирования	Слабая	Сильная

- Система управления базами данных:

Таблица 4. Система управления базами данных

Критерии	<i>MySQL</i>	<i>PostgreSQL</i>	<i>MongoDB</i>
Лояльность хостеров	Высокая	Низкая	Низкая
Наличие расширения в виде ГИС	Нет	Есть ( <i>PostGIS</i> )	Нет
Гибкость в плане хранения документов	Средняя	Средняя	Высокая
Модель организации баз данных	Реляционная	Реляционная	Документо-ориентированная
Формат по добавлению данных	<i>SQL</i>	<i>SQL</i>	<i>JSON</i>

- Система управления контентом для ГИС:

Так как ресурс – один из списка, рассмотрим его отдельно.

- *Mapbender* (рис. 13);
- *Поддержка разметок для вёрстки сайтов;*
- *Работа с определениями по формату YAML;*
- *Поддержка библиотек для вёрстки веб-ГИС: OpenLayers, jQuery, MapQuery;*
- *Работа с языком PHP;*
- *Возможность написания документации по ресурсу посредством Git.*

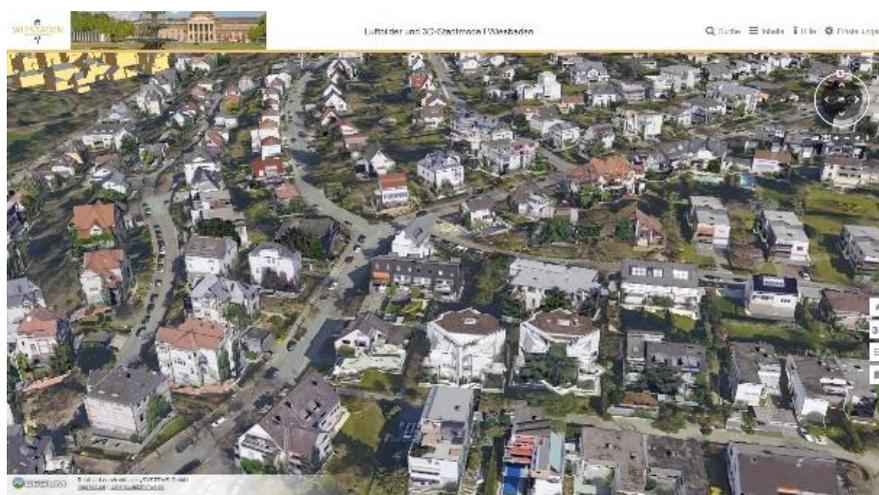


Рис. 13. Пример города Wiesbaden на Mapbender

### Заключение

Статистическая информация имеет огромное значение при принятии решений в экономике, проведении научных исследований и в образовании.

«Человек учится намного быстрее и лучше, если информация, необходимая для выполнения задачи, либо находится в готовом виде в окружающем мире, либо передается через ограничители» [1].

В итоге было проведены научно-исследовательские работы по статистическим информационным ресурсам, по принципам грамотного составления отчёта и по инструментам с технологиями ГИС.

### Литература

1. Норман Дональд А. Дизайн привычных вещей. – М.: Манн, 2018. С. 384.
2. Основные принципы официальной статистики. Принцип 1: Значимость, Объективность и Общедоступность [Электронный ресурс] – Электрон. текст. – Режим доступа: <https://unstats.un.org/unsd/methods/statorg/fp-russian.pdf>, свободный (дата обращения: 10.06.2020).
3. Федеральный закон от 29.11.2007 N 282-ФЗ (ред. от 18.04.2018) "Об официальном статистическом учете и системе государственной статистики в Российской Федерации", п. 10 ст. 5 [Электронный ресурс] – Электрон. текст. – Режим доступа: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_72844/bf4e92acc9726fed2ff772565df4d7c062e08967/](http://www.consultant.ru/document/cons_doc_LAW_72844/bf4e92acc9726fed2ff772565df4d7c062e08967/), свободный (дата обращения: 10.06.2020).
4. Берестова Т.Ф. Понятие «Информационные ресурсы» и другие компоненты теории информационного ресурсоведения // Научно-техническая информация. Серия 1. Организация и методика информационной работы, 2016, № 4 – УДК [001.102 : 002] – 027.21 : 001.102 – Электрон. текст. – Режим доступа: <http://lamb.viniti.ru/sid2/sid2free?sid2=J14398386>, свободный (дата обращения: 10.06.2020).
5. Stephen Few. Show me the numbers. PerceptualEdge, 2010. P. 174.
6. Stephen Few. Heatmaps: to Bin or Not to Bin? // Visual Business Intelligence Newsletter. October/November/December 2017.